



**University of  
Zurich<sup>UZH</sup>**

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2012

---

## **Rapid Synthesis Of Defined Eukaryotic Promoter Libraries**

Rajkumar, Arun S ; Maerkl, Sebastian J

**Abstract:** Current gene synthesis methods allow the generation of long segments of dsDNA. We show that these techniques can be used to create synthetic regulatory elements and describe a method for the creation of completely defined, synthetic variants of the PHO5 promoter from the budding yeast *Saccharomyces cerevisiae*. 128 promoters were assembled by high-temperature ligation, cloned into plasmids by isothermal assembly, maintained in *E. coli*, and consequently transformed into yeast by homologous recombination. Synthesis errors occurred at frequencies comparable to, or lower than those achieved with current gene synthesis methods. The promoter synthesis method reported here is robust, fast, and readily accessible. Synthetically engineered promoter libraries will be useful tools for dissecting the intricacies of promoter input-output functions, and may serve as tunable components for synthetic genetic networks.

DOI: <https://doi.org/10.1021/sb300045j>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-78704>

Journal Article

Originally published at:

Rajkumar, Arun S; Maerkl, Sebastian J (2012). Rapid Synthesis Of Defined Eukaryotic Promoter Libraries. *ACS Synthetic Biology*, 1:483-490.

DOI: <https://doi.org/10.1021/sb300045j>

# Rapid Synthesis Of Defined Eukaryotic Promoter Libraries

Arun S. Rajkumar and Sebastian J. Maerkl\*.

School of Engineering, Institute of Bioengineering, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland.

\*To whom correspondence should be addressed. Tel: +41 21 6937835; Email: sebastian.maerkl@epfl.ch

## ABSTRACT

Current gene synthesis methods allow the generation of long segments of dsDNA. We show that these techniques can be used to create synthetic regulatory elements and describe a method for the creation of completely defined, synthetic variants of the *PHO5* promoter from the budding yeast *Saccharomyces cerevisiae*. 128 promoters were assembled by high-temperature ligation, cloned into plasmids by isothermal assembly, maintained in *E. coli*, and consequently transformed into yeast by homologous recombination. Synthesis errors occurred at frequencies comparable to, or lower than those achieved with current gene synthesis methods. The promoter synthesis method reported here is robust, fast, and readily accessible. Synthetically engineered promoter libraries will be useful tools for dissecting the intricacies of promoter input-output functions, and may serve as tunable components for synthetic genetic networks.

## INTRODUCTION

Gene synthesis, or more broadly, the synthesis of long dsDNA from smaller ssDNA components has recently become an important tool in synthetic biology, genetic, and genome engineering (1). With respect to efficiency and fidelity, these methods work best for the assembly of segments or subassemblies of DNA roughly 1kb or less in size (2). When parallelized and coupled with other assembly techniques, it is possible to build entire genomes (3), or carry out high-throughput gene synthesis (4-6). Such feats have been made possible by the maturation of gene synthesis protocols, which normally use ligation or polymerase-based methods for assembly (2,7). By careful design of components and the use of high-fidelity or mismatch-cleaving enzymes, error rates of ~0.1% (or 1 error per kb) could be achieved (2).

Libraries of native promoters can yield useful information on the rules governing gene regulation, an approach that has recently been used for yeast ribosomal protein genes (8). Random assembly of promoter components has also been used to study promoter architecture. However, the scope of such studies is limited by the inherent randomness of assembled promoters, allowing only broad inferences on promoter architecture and gene

regulation to be derived (9,10). Such an approach has nonetheless revealed the role of low-affinity TF-DNA interactions in gene regulation (11), the modularity of core promoter elements (12), and demonstrated the ability of TFs to switch function depending on environmental conditions (13). But randomly synthesized promoters can introduce levels of complexity that make it difficult to interpret the resulting data.

Defined synthetic promoter libraries allow the study of gene regulation in a systematic manner. By engineering promoters it is possible to systematically investigate how individual regulatory elements contribute to the behavior of the promoter as a whole. A defined synthetic promoter can be modular, allowing the insertion, duplication, removal, or displacement of regulatory elements with no loss of combinatorial flexibility. More specifically, modifying the context of a given regulatory element, its initial accessibility to binding proteins, or its distance from the transcription start site can provide insight into rules governing promoter architecture. While large libraries of modified short promoters can be directly created by oligonucleotide synthesis or PCR(14,15), similar-sized libraries of entire eukaryotic promoters require different protocols that can be efficiently integrated into large-scale workflows.

We have developed a method requiring neither specialized equipment nor reagents for generating large libraries of defined eukaryotic promoters and generated a total of 128 promoter variants modifying both Pho4 binding sites in the *PHO5* promoter. The entire workflow requires 13-16 days for one batch of promoters to be chromosomally integrated and fully sequence verified. When induced, promoters from the library were found to be fully functional. The cost per promoter variant compares favorably to commercial site-directed mutagenesis kits, and the cost of individual promoters decreases with increased library sizes.

## Results And Discussion

### *Construction of synthetic promoters*

We tiled the native *PHO5* promoter into overlapping oligo pairs to create a modular synthetic promoter allowing individual regulatory components to be independently assembled (Figure 1, Table S1). We chose 90bp long oligos to minimize synthesis costs. Given that the promoter has high AT content, it was also necessary to have sufficiently long overlaps allowing them to anneal at high temperatures. While constructing the exposed-site variant library, we found it necessary to change the length of the left arm to accommodate site variants and minimize oligo synthesis expenses. In all cases, the overlaps between oligo pairs had melting points of at least 58°C (Table S2), which set a lower limit to the ligation temperature.

We found high-temperature, single-step ligation to be a simple yet robust promoter assembly method. While a 4-hour ligation was sufficient to yield assembled promoter upon amplification, ligation was normally allowed to run overnight to maximize yield. Our choice of ligase was based on the fact that 9°N ligase preferentially ligates long overlaps (12 bases; New England Biolabs, personal communication). With the reaction

conditions described here we were consistently able to amplify 1.5-2pmol of full-length promoter. As our synthetic promoters are derived from a native promoter, constraints on the assembly scheme imposed by promoter sequence made us forego normalization of the melting temperature of overlaps between segments, which is a common used to improve assembly efficiency. However, promoter synthesis was efficient and robust without this precaution.

We maintained each library in *E. coli* to ensure a stable copy of each construct from which DNA for transformation into yeast could be readily generated. Gibson assembly consistently yielded sufficient quantities of plasmid for transformation (16). Changing neither the promoter-to-plasmid backbone ratio nor assembly time significantly improved the yield or quality of assembled plasmid. In contrast to ligation, Gibson assembly carries little to no risk of the plasmid backbone re-ligating and contaminating the transformants, in addition to allowing a scarless fusion of promoter and reporter sequences. Finally, the exonuclease used in a Gibson assembly digests any secondary ligation products, thus eliminating the need for gel purification of full-length promoter assemblies prior to cloning.

#### *Fidelity of promoter assembly*

It was our intention to create a library of *PHO5* promoters containing all possible variants of the nucleosomal and exposed Pho4 E-box half-sites. We constructed 128 independently varied half-sites, and maintained them in *E. coli* and budding yeast. At each step, we verified the library sequences to determine the presence and frequency of assembly errors. We considered a promoter to be functional if it contained no more than 1% errors (indels, mutations and ambiguous sequencing base-calls) in regions outside known regulatory elements. The quality and coverage of our promoter libraries are listed in Table 1. The absence of errors in identical locations in each suggested that no errors arose from oligonucleotide synthesis.

#### *E. coli library quality*

We successfully cloned all variants of each Pho4 site into *E.coli*, with each promoter having at least two-fold coverage. All promoters had ~1.6 errors per kilobase and 84% of promoters (108 out of 128) were perfect (Table 2, Figures 2a and 2c). The remaining promoters had mostly single-base deletions. However, we were able to retrieve clones we considered to be functional for each imperfect promoter variant. The majority of promoter synthesis errors were deletions, with single insertions, mutations and ambiguous base-calls from sequencing making up the remainder of errors. All sequence errors were distributed randomly across the promoters (Figures 3a and 3c).

#### *Yeast library quality*

124 out of 128 synthetic promoters, amounting to 62 variants of each Pho4 site, were transformed into yeast by homologous recombination. While the average error rate was lower than that of the *E. coli* library, the errors were spread out over more promoters. Nonetheless, over 50% and 70% of nucleosomal- and exposed-site promoters, respectively, were error-free (Figures 2b and 2d). Errors in nucleosomal-site promoters were mostly ambiguous base-calls, whereas those in exposed-site promoters were single-

base deletions. Unlike the bacterial library, the majority of sequence errors in the yeast library were ambiguous base-calls from sequencing reads, followed by single-base deletions. Most of these ambiguous base-calls were located either in or near repetitive stretches of the promoter or near the beginning of the sequencing read, while other errors were randomly distributed over the promoter (Figures 3d and 3c).

#### *In vivo functionality of synthetic promoters*

To test the functionality of our promoter library we measured the induction of 12 synthetic promoters whose modified nucleosomal and exposed sites cover the entire range of Pho4 affinities (Figure 4). It has previously been shown that the induction kinetics of Pho4-regulated promoters during Pi starvation are well described by a time-dependent Hill function (17). This was also the case for our synthetic promoters (Table S3). As expected,  $F_{max}$  the Hill fit parameter related to the final induction level was dependent on the affinity of the modified nucleosomal and exposed Pho4 sites (Figures 4b and 4e). On the other hand the time to half-maximal induction  $t_{1/2}$  (used as a measure of the time to induction), only depended upon the affinity of the exposed Pho4 site (Figure 4f).

*De novo* synthesis of gene-sized dsDNA is a powerful tool in synthetic biology and genomics research. The method described here allows the rapid creation of defined synthetic promoters, which in turn permit the systematic exploration of the structure-function relationship of eukaryotic promoters. We have taken an existing, well-studied promoter from yeast and developed a workflow to create libraries of defined variants, using a straightforward but effective promoter synthesis protocol. To ensure that promoter assembly and amplification was carried out at the highest fidelity possible, we used a thermostable ligase that would not ligate short mis-annealed overlaps and a polymerase with low error rate. When compared to existing gene synthesis methods, our method yielded error rates comparable to the best of these methods (4,18-24, Table S4). While single-step ligation is one of the oldest assembly methods (25), our work shows that it can assemble DNA with a fidelity matching state-of-the-art techniques. As most errors encountered seem to appear during transformation into *E. coli*, it is unclear whether post-assembly error correction would yield better results.

To identify the origin of sequence errors found after cloning we asked whether single-base errors occurred at breakpoints between component oligos. We measured the distance of deletions in the bacterial library from the nearest breakpoints and found that they were no closer to breakpoints than randomly chosen locations on the promoter (Figure S2). It is possible, however that errors cluster near short component oligos with lower annealing temperatures as may be the case for the oligo pair containing the exchangeable exposed site. The preponderance of deletions in gene synthesis errors could be attributed to the use of a proofreading polymerase, as has previously been reported (23).

A time estimate for creating a synthetic promoter library consisting of 50 members is provided in Table S5. Cloning into *E. coli* did not significantly increase the time taken to construct the library, as the rate-limiting step is the integration of the library into yeast by

homologous recombination. If used in conjunction with robotic handling, the time to create a library could be considerably reduced.

While best suited for studying derivatives of a native promoter, synthetic promoter library construction offers a number of advantages over site-directed mutagenesis. The latter typically offers the ability to modify a few bases, and only a few commercially available kits allow the simultaneous modification of several bases. Creating a synthetic promoter can introduce as many sequence modifications as needed with high efficiency. Table S6 provides the cost of creating a single promoter from our library and maintaining it in *E.coli*, and compares it against the price for a single mutagenesis reaction from a set of commercially available site-directed mutagenesis kits. The comparison shows that the cost of one such reaction (36.03CHF) compares favorably with the price range of commercial mutagenesis kits (32-50CHF not counting primer and sequencing costs).

We presented the induction kinetics of a small subset of our promoter library to demonstrate its functionality. Our results on induction timing and levels on these specific variants of the *PHO5* recapitulate earlier findings (17), they also reveal subtleties in the role of each Pho4 binding site in transcription regulation. While changing the affinity of either site in the promoter affects the level of induction, the dynamic range of the induction levels is larger for the exposed site. The effect of exposed site affinity on the induction kinetics is also non-trivial, and warrants further investigation.

A promoter is expected to be functional even if it contains a few errors, so long as these errors lie outside of regulatory elements. Induction experiments on perfect and error-containing clones of the same promoter found no difference in induction kinetics (Figure 5a). Indeed, deletions as large as 4 bases in a non-regulatory region have no effect on a promoter's output (Figure 5b). On the other hand, we found that errors in or near known or predicted regulatory elements do affect promoter induction characteristics and can do so with unexpected results. The decrease in induction for promoter D3E (Figure 5c) can be attributed to deletions in a known binding region for the co-regulator Pho2 (26) and a predicted Swi5 binding site (27). The effect of defects in G4N is harder to explain as the insertion near the TATA box is expected to neither improve nor weaken it with regard to the consensus sequence (28). This mutation nonetheless increased induction by nearly 2-fold (Figure 5d).

In summary, we developed a robust method for generating synthetic promoter libraries, and demonstrated their utility as tools for studying gene regulation. The ability to quickly and robustly generate hundreds of eukaryotic promoter variants, integrate them into the genome, and measure their output using a reporter gene will be useful in deciphering the rules governing gene regulation. Finally, defined and well-characterized promoter libraries will be valuable components for building synthetic genetic networks.

## Methods

### *Synthetic promoter design*

Our promoter library consists of defined synthetic derivatives of the yeast *PHO5* promoter. The *PHO5* promoter regulates the synthesis of an acid phosphatase during inorganic phosphate (Pi) starvation and is one of the best-studied eukaryotic promoters. This promoter is regulated by the transcription factor Pho4, which is thought to bind to two sites (Figure 1a and Figure S1): an initially exposed, low-affinity site, and a high-affinity site covered by nucleosomes (29,30). The presence of exposed and nucleosomal Pho4 sites allows independent control of *PHO5* induction thresholds and expression levels (17).

Each synthetic promoter was designed to consist of interchangeable pieces containing the nucleosomal and exposed Pho4 sites flanked by two constant arms (Figure 1a). Pho4 binding sites are 10-mers centered on a symmetric, hexameric E-box (31). The library described in this paper consisted of variants of either site's E-box (Figure 1b).

The promoter was taken to be the 800 bp sequence upstream of the *PHO5* open reading frame, (chrII:430946-431745, minus strand) and the sequence was taken from the June 2008 build of the *S. cerevisiae* genome available at the UCSC Genome Browser (<http://genome.ucsc.edu/>). We tiled the native promoter into 90bp-oligonucleotide pairs (Table S1) with 5' and 3' overhangs 30bp long (Figure 1a). We normalized the overlap between oligo pairs with regards to length rather than melting temperature due to constraints imposed by the promoter. Melting temperatures of the overlaps were calculated using the Oligo Analysis Tools found on the Operon website (<http://www.operon.com/technical/toolkit.aspx>), and each oligo was checked for secondary structure at the ligation temperature using the mFold server (32). The oligos making up the constant arms were ordered from IDT (Leuven, Belgium) and the Pho4 site variant oligos were ordered from Invitrogen (Carlsbad, US).

#### *Promoter assembly and amplification*

300pmol aliquots of each oligo pair were phosphorylated overnight at 37°C using 10U T4 polynucleotide kinase (NEB) in the provided buffer (70mM Tris-HCl, 10mM MgCl<sub>2</sub>, 5mM DTT at pH 7.6) supplemented with 1mM ATP, followed by heat inactivation for 20 minutes at 65°C. The phosphorylated oligos were directly used for promoter assembly without further purification. For assembly, 3pmol of each oligo pair forming the flanking arms was mixed with an equimolar amount of oligo pairs containing the Pho4 site variant and a one-pot ligation was carried out in a volume of 15μL using 9°N ligase (NEB), with a final oligo concentration of 200nM in 1x ligase buffer (10mM Tris-HCl, 0.6mM ATP, 2.5mM MgCl<sub>2</sub>, 2.5mM DTT, 0.1% Triton X-100 at pH 7.5). The mixture was heated to 95°C for 6 minutes, cooled to 60°C at a rate of 0.1°C per minute and incubated overnight (typically 14h) at 60°C. 9°N ligase was chosen for its ability to preferentially ligate long overlaps and thus avoid mis-ligation. The ligation temperature was chosen to be close to, but lower than, the average melting point of the overlaps between oligo pairs (63°C, Table S2). 4fmol of the assembled promoter was used as a template for amplification. A typical PCR reaction was carried out with Phusion 2x Master Mix with HF buffer (Finnzymes) and 200nM of each primer in a volume of 40μL. The forward primer incorporates a 30bp synthetic sequence upstream of the promoter to label the promoter as 'synthetic' and to track its integration in yeast. The promoter was amplified using 30

cycles of 10s at 98°C, 20s at 65°C, and 15s at 72°C, followed by a final extension at 72°C for 6min (Figure 6a).

#### *Plasmid assembly and cloning.*

Each promoter was cloned into plasmid pBS34 (obtained from the Yeast Resource Center, University of Washington) directly upstream of the mCherry gene (Figure 1C) using one-step, isothermal Gibson assembly (16). We chose mCherry as our reporter because of its short maturation time (33) and the low auto-fluorescence of yeast in its emission range. For the purpose of isothermal assembly, each promoter has the first 25 bases of the mCherry gene added to its 3' end during amplification, and the linearized plasmid backbone in turn bears the promoter's 30bp synthetic tag. In brief, 5μL of backbone and promoter were added to 15μL of Gibson reaction mixture containing 0.75U/mL T5 exonuclease, 25U/mL Phusion polymerase and 4U/mL Taq DNA ligase in 100mM Tris-HCl, 5mM MgCl<sub>2</sub>, 10mM DTT, 1mM NAD, 5% PEG, and 200μM of each dNTP. The reactions were carried out for 1hr using a promoter-to-backbone ratio of 2:1 in a volume of 15μL at 50°C (Figure 6b). 2μL of the assembly mix was directly transformed into DH5α *E. coli* cells made competent by the CaCl<sub>2</sub> method.

Three clones from each transformation were screened for the promoter-containing plasmid by colony PCR. Colony PCR was carried out using primers flanking mCherry in the pBS34 backbone; the presence of a promoter-containing plasmid would yield a product ~1.6kb, roughly twice the size of mCherry (Figure 6c). Once transformed, the promoter library was sent for sequencing using a sequencing primer 72bp downstream of the mCherry start codon.

#### *Library transformation into yeast*

Promoter constructs with mCherry and the kanMX6 marker from the plasmid were directly amplified from each clone's glycerol stock for transformation into yeast by homologous recombination at the LYS2 locus (Figure 6d). 40bp sequence insertion tags were added during amplification. The amplified constructs were transformed into yeast strain BY4741 (MATa *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) by the lithium acetate/PEG method (34). Colonies resistant to 350μg/mL G418 were screened for lysine auxotrophy to verify integration at the LYS2 locus. We picked up to three colonies per strain for final confirmation by colony PCR.

We used semi-nested colony PCR to confirm correct insertion of the synthetic promoter at the LYS2 locus by using two forward primers, one targeting ~100bp upstream of the insertion site and the second targeting the synthetic tag (Figure 1d). Clones with both positive colony PCR products (Figure 2e, lanes 3-5) were archived and their synthetic promoters re-amplified from yeast for sequence verification (Figure 2f). All primers used in promoter assembly, amplification, and colony PCR are listed in Table S7.

#### *Predicting Pho4 affinity to engineered regulatory elements*

We calculated the probability of Pho4 binding to a 24bp region centered on the Pho4 E-box using a simple *in silico* model. Binding probabilities ( $P_{occ}$ ) were calculated from measured Pho4 binding energies to sites on a 12bp long sliding window (31,35).



### *Kinetic induction measurements*

Strains were grown in YPD supplemented with 10mM Pi and 200µg/µL G418 at 30°C for 26 hours, then diluted 30-fold in synthetic complete medium with 10mM Pi and allowed to re-enter log phase. Cells were washed twice in Pi-free synthetic medium and diluted to a starting OD of 0.1-0.2 in Pi-free medium. mCherry fluorescence (587nm excitation and 610nm emission, 9nm bandwidth) was measured every 6 minutes for 16 hours on a plate reader (BioTek SynergyMx) and normalized to cell number by dividing by the optical density. Time-dependent Hill functions were fit to the normalized induction curve.

### **Author Contributions**

S. J. M and A. S. R designed the library content; A.S.R. designed the synthetic promoter structure and carried out the experiments and data analysis; both authors wrote the manuscript.

### **Supporting Information Available**

Supplementary Tables 1-7, Supplementary Figures 1 and 2 and Supplementary References. This information is available free of charge via the Internet at <http://pubs.acs.org/>.

### **Acknowledgements**

We wish to thank the Shore lab at the University of Geneva for their assistance. This work was supported by a SystemsX.ch grant DynamiX-RTD (2008/005); and the École Polytechnique Fédérale de Lausanne.

### **References**

1. Hughes, R. E., Miklos, A. E. and Ellington, A. E. (2011) Gene synthesis: Methods and applications. *Methods Enzymol.*, 498, 277-309.
2. Czar, M. J, Anderson, J. C., Bader, J. S. and Peccoud, J. (2009) Gene synthesis demystified. *Trends Biotechnol.*, 27, 63-72.
3. Smith, H. O., Hutchison III, C. A., Pfannkoch, C. and Venter, J. C. (2003) Generating a synthetic genome by whole genome assembly: ΦX174 bacteriophage from synthetic oligonucleotides. *Proc. Natl. Acad. Sci U.S.A.*, 100, 15440-15445.
4. Quan, J., Saeem, I., Tang, N., Ma, S., Negre, N., Gong, H., White, K. P. and Tian, J. (2011) Parallel on-chip gene synthesis and application to optimization of protein expression. *Nat. Biotechnol.*, 29, 449-453.
5. Cox, J. C., Lape, J., Sayed, M. A. and Hellinga, H. W. (2007) Protein fabrication automation. *Protein Sci.*, 16, 379-390.
6. Tian, J., Gong, H., Sheng, N., Zhou, X., Gulari, E., Gao, X. and Church, G. (2004) Accurate multiplex gene synthesis from programmable DNA microchips. *Nature*, 432, 1050-1054.

7. Xiong, A.-S., Peng, R.-H., Zhuang, J., Liu, J.-G., Gao, F., Chen, J.-M., Cheng, Z.-M. and Yao, Q.-H. (2008) Non-polymerase-cycling-assembly-based chemical gene synthesis: Strategies, methods, and progress. *Biotechnol. Adv.*, *26*, 121-134.
8. Zeevi, D., Sharon, E., Lotan-Pompan, M., Lubling, Y., Shipony, Z., Raveh-Sadka, T., Keren, L., Levo, M., Weinberger, A. and Segal, E. (2011) Compensation for differences in gene copy number among yeast ribosomal proteins is encoded within their promoters. *Genome Res.*, *21*, 2114-2128.
9. Ligr, M., Siddhartan, R., Cross, F. R. and Siggia, E. D. (2006) Gene expression from random libraries of yeast promoters. *Genetics*, *172*, 2113-2122.
10. Cox III, R. S., Surette, M. G. and Elowitz, M. B. (2007) Programming gene expression with combinatorial promoters. *Mol. Syst. Biol.*, *3*, 145.
11. Gertz, J., Siggia, E. D. and Cohen, B. A. (2009) Analysis of combinatorial cis-regulation in synthetic and genomic promoters. *Nature*, *457*, 215-218.
12. Mogno, I., Vallania, F., Mitra, R. D. and Cohen, B. A. (2010) TATA is a modular component of synthetic promoters. *Genome Res.*, *20*, 1391-1397.
13. Gertz, J. and Cohen, B. A. (2009) Environment-specific combinatorial cis-regulation in synthetic promoters. *Mol. Syst. Biol.*, *5*, 244.
14. Patwardhan, R. P., Lee, C., Litvin, O., Young, D. L., Pe'er, D. and Shendure, J. (2009) High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nature Biotechnol.*, *27*, 1173-1175.
15. Schlabach, M. R., Hu, J. K., Li, M. and Elledge, S. J. (2010) Synthetic design of strong promoters. *Proc. Natl. Acad. Sci. U.S.A.*, *107*, 2538-2543.
16. Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison III, C. A., and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods*, *6*, 343-347.
17. Lam, F. H., Steger, D. J., and O'Shea, E. K. (2008) Chromatin decouples promoter threshold from dynamic range. *Nature*, *453*, 246-251.
18. Bang, D. and Church, G. M. (2008) Gene synthesis by circular assembly amplification. *Nat. Methods*, *5*, 37-39.
19. Borovkov, A. Y., Loskutov, A. V., Robida, M. D., Day, K. M., Cano, J. A., Le Olson, T., Patel, H., Brown, K., Hunter, P. D and Sykes, K. F. (2010) High-quality gene assembly directly from unpurified mixtures of microarray-synthesized oligonucleotides. *Nucleic Acids Res.*, *38*, e180.
20. Xiong, A.-S., Yao, Q.-H., Peng, R.-H., Li, X., Fan, H.-Q., Cheng, Z.-M. and Yi, L. (2004) A simple, rapid, high-fidelity and cost-effective PCR-based two-step DNA synthesis method for long gene sequences. *Nucleic Acids Res.*, *32*, e98.
21. Gao, X., Yao, P., Keith, A., Ragan, T. J. and Harris, T. K. (2003) Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis: a novel method of primer design for high-fidelity assembly of longer gene sequences. *Nucleic Acids Res.*, *31*, e143.
22. Young, L., and Dong, Q. (2004) Two-step total gene synthesis method. *Nucleic Acids Res.*, *32*, e59.
23. Wu, G., Wolf, J. B., Ibrahim, A. F., Vadasz, S., Gunasinghe, M. and Freeland, S. J. (2006) Simplified gene synthesis: A one-step approach to PCR-based gene construction. *J. Biotech.* *124*, 496-503.

24. Gordeeva, T. L., Borschevskaya, L. N. and Sineoky, S. P. (2010) Improved PCR-based gene synthesis method and its application to the *Citrobacter freundii* phytase gene codon modification. *J. Microbiol. Methods*, 81, 147-152.
25. Ecker, D. J., Khan, M. I., Marsh, J., Butt, T. R. and Cooke, S. T. (1987) Chemical synthesis and expression of a cassette-adapted ubiquitin gene. *J. Biol. Chem.*, 262, 3524-3527.
26. Barbaric, S., Münsterkötter, M., Coding, C. and Hörz, W. (1998) Cooperative Pho2-Pho4 interactions at the *PHO5* promoter are critical for binding of Pho4 to UASp1 and for efficient transactivation by Pho4 at UASp2. *Mol. Cell. Biol.*, 18, 2629-2639.
27. MacIsaac, K. D., Wang, T., Gordon, D. B., Gifford, D. K., Stormo, G. D. and Fraenkel, E. (2006) An improved map of conserved regulatory sites for *Saccharomyces cerevisiae*. *BMC Bioinformatics*, 7, 113.
28. Basehoar, A. D., Zanton, S. J. and Pugh, J. F. (2004) Identification and distinct regulation of yeast TATA box-containing genes. *Cell*, 116, 699-709.
29. Vogel, K., Hörz, W. and Hinnen, A. (1989) The two positively acting regulatory proteins Pho2 and Pho4 physically interact with *PHO5* upstream activation regions. *Mol. Cell. Biol.*, 9, 2050-2057.
30. Svaren, J. and Hörz, W. (1997) Transcription factors vs nucleosomes: regulation of the *PHO5* promoter in yeast. *Trends Biochem. Sci.*, 22, 93-97.
31. Maerkl, S. J. and Quake, S. R. (2007) A systems approach to measuring the binding energy landscapes of transcription factors. *Science*, 315, 233-237.
32. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, 31, 3406-3415.
33. Shaner, N. C., Campbell, R. E., Steinbach, P. A., Niepmans, B. N. G., Palmer, A. E. and Tsien, R. Y. (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat. Biotechnol.*, 22, 1567-1572.
34. Gietz, R. D. and Schiestl, R. H. (2007) High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.*, 2, 31-34.
35. Granek, J. A. and Clarke, N. D. (2005) Explicit equilibrium modeling of transcription-factor binding and gene regulation. *Genome Biol.*, 6, R87.

## TABLES AND FIGURE LEGENDS

**Table 1.** Promoter library

<i>PHO5</i> promoter site modified	Host	promoters with the correct site	clones with the correct site	perfect promoters	clones with perfect promoter
Nucleosomal	<i>E.coli</i>	64	152	54	86
	yeast	62	84	37	44
Exposed	<i>E.coli</i>	64	155	54	91
	yeast	62	93	45	67

**Table 2.** Promoter library sequence errors

Modified Pho4 site	Host	Bases sequenced	Deletions per kb	Insertions per kb	Mutations per kb	Ambiguous base calls per kb	Total errors per kb
Nucleosomal	<i>E.coli</i>	123866	1.17	0.09	0.073	0.27	1.6
	yeast	67192	0.104	0.045	0.104	0.67	0.91
Exposed	<i>E.coli</i>	121399	1.64	0.016	0.066	0.01	1.67
	yeast	75170	0.2	0.027	0.013	0.17	0.41
<b>Both</b>	<i>E.coli</i>	<b>245265</b>					<b>1.64</b>
	yeast	<b>142362</b>					<b>0.65</b>

**Figure 1:** Promoter synthesis and library generation. (a) The *PHO5* promoter sequence is tiled into 9 component oligo pairs, keeping the Pho4 sites ('Exp' and 'Nuc') on separate pairs. Each promoter variant is assembled from phosphorylated oligo pairs as described in Materials and Methods. (b) Each promoter library is ligated at 60°C, then amplified for cloning. (c) Each promoter variant is cloned into a plasmid containing mCherry and a yeast selection marker using Gibson assembly. Successful transformants are confirmed by colony PCR and their promoter sequences are subsequently confirmed. (d) Promoters with a correct sequence, along with the reporter and marker, are amplified from the plasmid and transformed into yeast. G418-resistant clones exhibiting lysine auxotrophy are checked for correct integration of the construct by nested colony PCR. The products of primers F1 and R1 confirm the presence of the synthetic promoter whereas the products of F2 and R1 confirm its integration. Clones giving positive products for both primer pairs have their entire promoters amplified using primers F2 and R2 for sequence verification. The entire assembly process from oligo pairs to yeast strain takes 10-14 days for a batch of promoters.

**Figure 2:** Synthetic promoter library quality. The promoter library charts are sorted vertically according by modified site (a,c) and horizontally by host (b,d). The first pie chart in each panel shows the number of perfect promoters per variant, the exploded slice shows the errors of the imperfect promoters, and the third chart shows the error

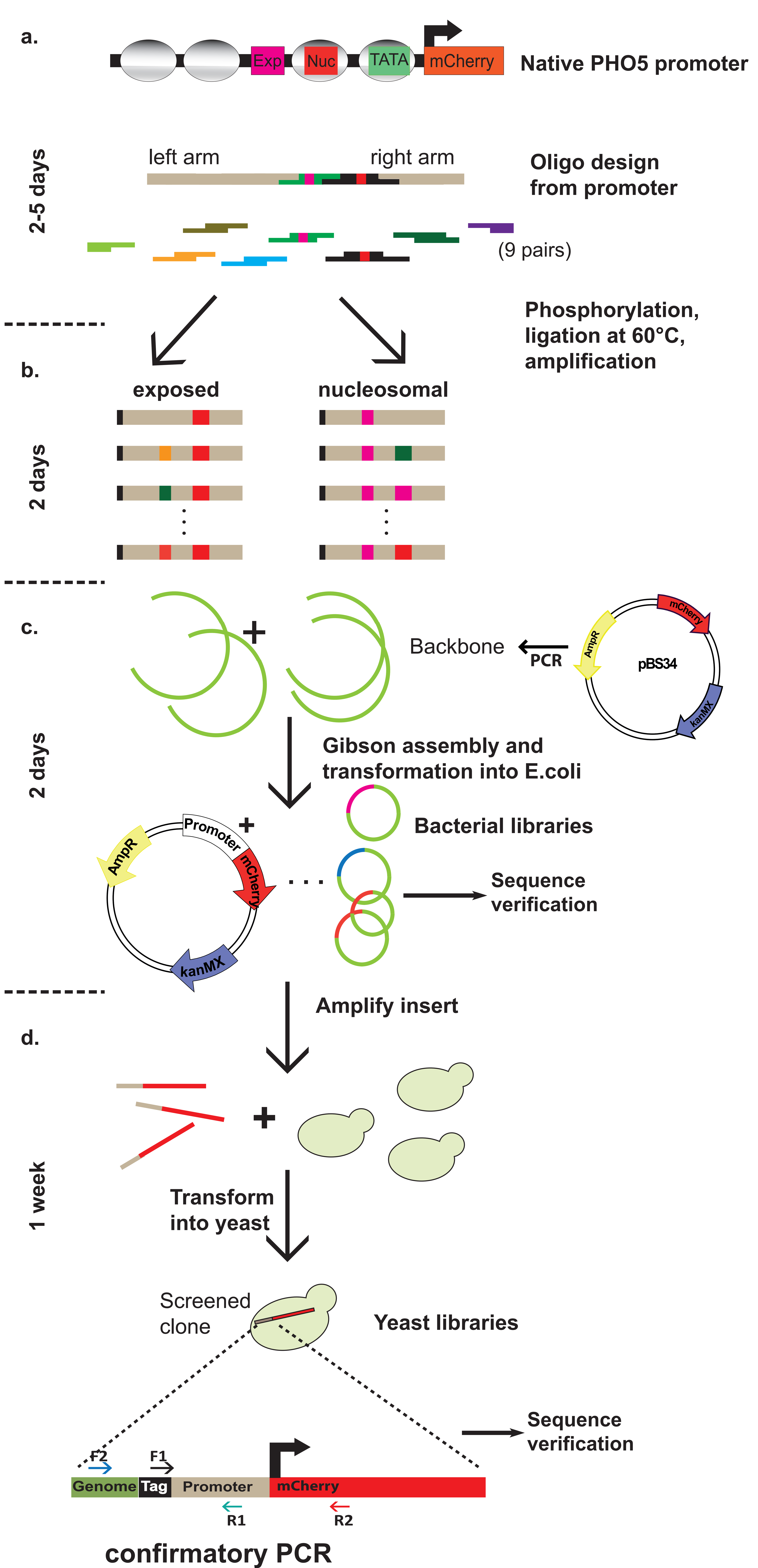
distribution for all clones with sequence information. Further details can be found in Tables 1 and 2.

**Figure 3:** Locations of sequence errors for nucleosomal-site (a,b) and exposed-site (c,d) synthetic promoters. 'Other deletions' include any deletion longer than one base. The major errors are single deletions distributed randomly over the promoter.

**Figure 4:** Characterization of synthetic promoter functionality. Induction curves for six variants of the nucleosomal (a) and exposed (d) Pho4 sites (including the wild-type *PHO5* promoter) under phosphate starvation. (b,e) Effect of modified site affinity on induction level, estimated from  $F_{max}$  from the Hill fit. (c,f) Effect of site affinity on time to half-maximal induction  $t_{1/2}$ .  $F_{max}$  and  $t_{1/2}$  values are the average of 5-7 independent experiments for each promoter.

**Figure 5:** The effect of sequence errors on promoter activity. (a,b) The presence of deletions or insertions in non-regulatory regions in a promoter has no effect on induction behavior. (c-d) Errors in or near known regulatory regions do affect promoter activity. (c) Deletions in a Pho2 binding site decrease induction levels significantly and an insertion near the TATA box (d) increases induction. Promoter maps below each graph indicate regulatory sites of interest and the presence of errors in defective promoters marked by 'x'.

**Figure 6:** Representative synthetic promoter products at each step of library assembly. (a) Amplification of eight synthetic promoters with modified nucleosomal sites. The added synthetic tag and overlap sequences for Gibson assembly increase the product's length from 800bp to over 900bp. (b) Time course of a typical Gibson assembly of promoter (Pr) and backbone (Ba) to product (P). (c) Colony PCR of promoters transformed into *E.coli*. The PCR product corresponds to the promoter-mCherry construct. (d) Direct amplification of promoter-mCherry-kanMX cassettes from glycerol stocks. (e) Colony PCR of G418-resistant, *Lys*<sup>-</sup> colonies transformed into yeast. A successful nested PCR yields two products positive for the correct synthetic promoter (T) and integration (L). (f) Direct amplification of the synthetic promoter for sequence verification from yeast. PCR-positive colonies from panel (E) were selected for this final amplification. Markers used in the gels were either the Fermentas GeneRuler 100bp (M) or 1kb (M2) ladder.

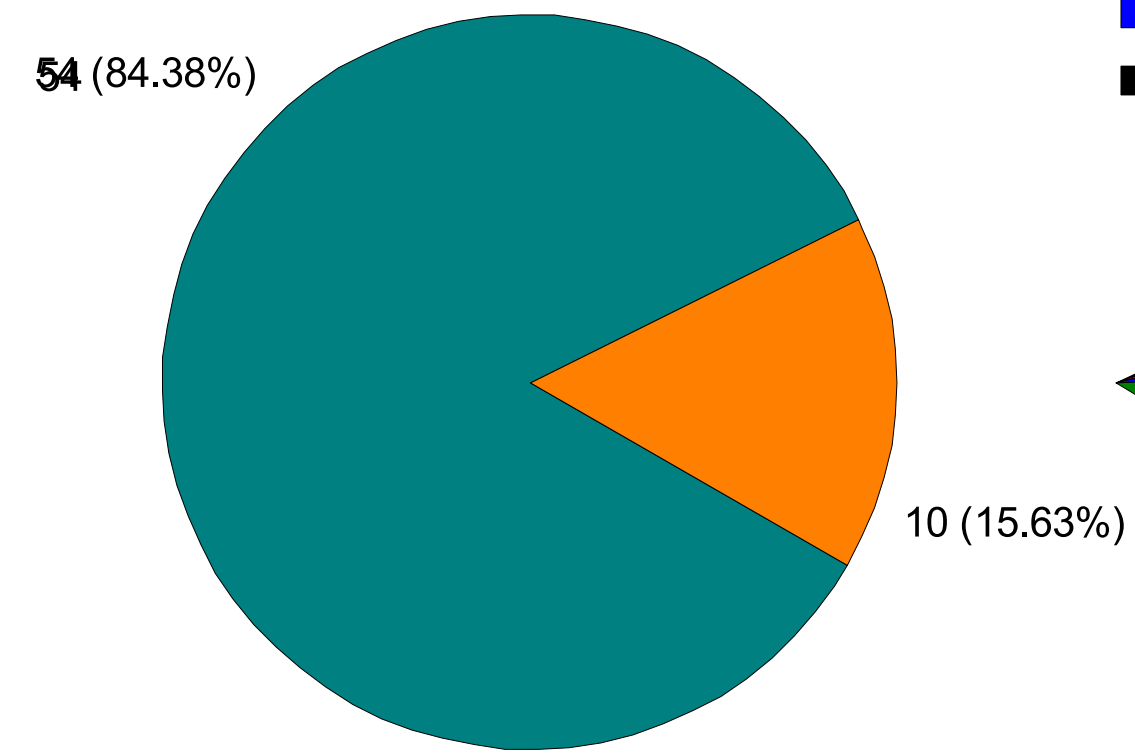




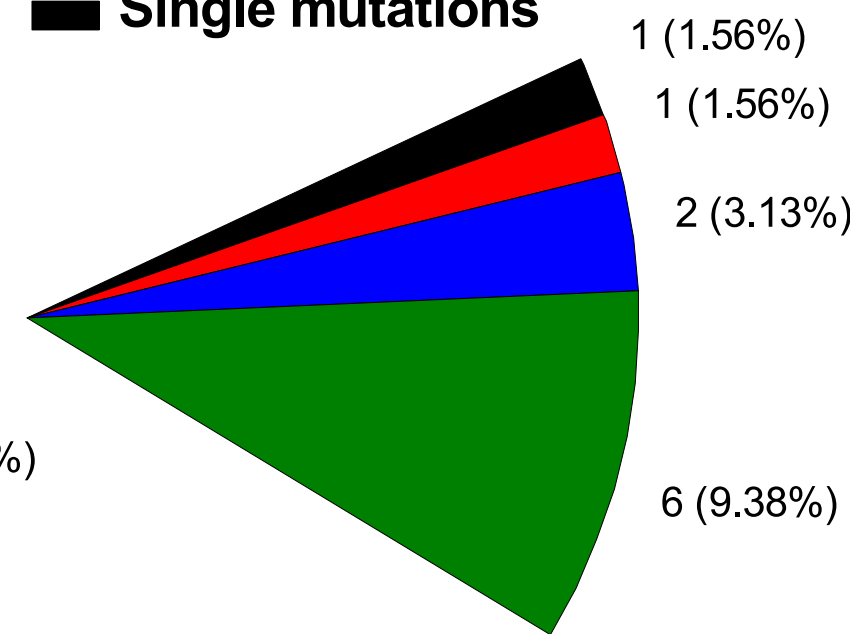
Nucleosomal site

a.

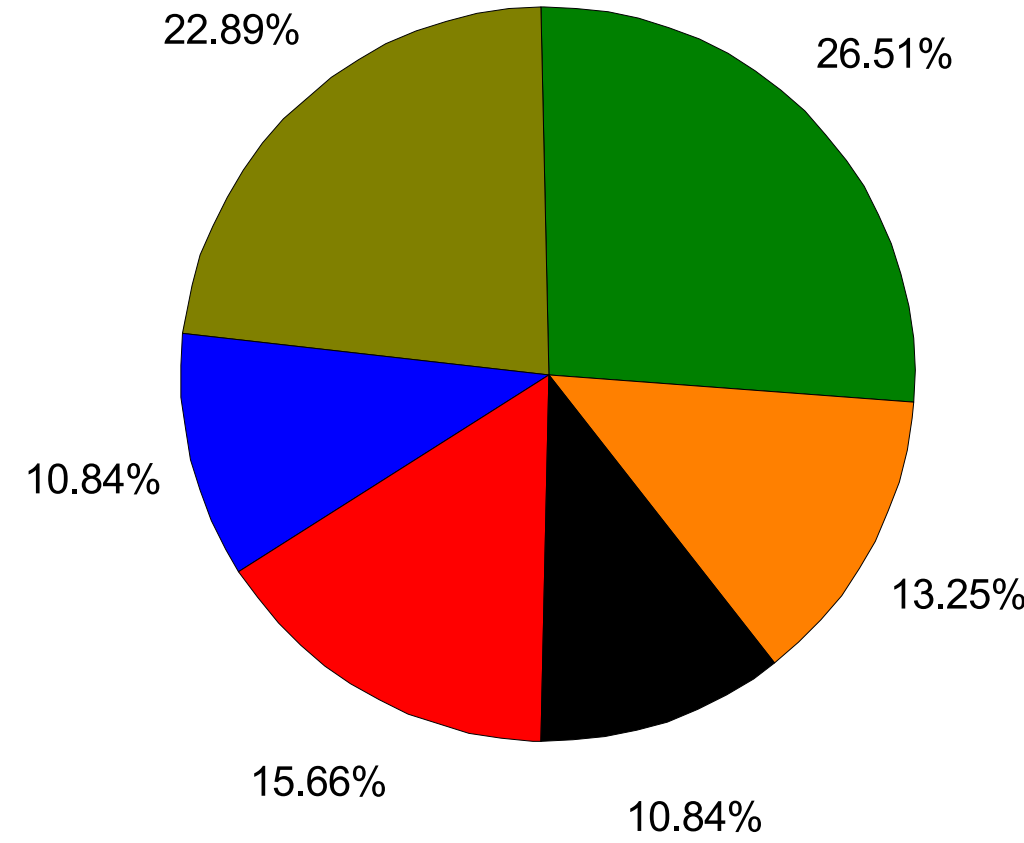
Perfect promoters  
Promoters with errors



Single deletions  
Short deletions (2-4 bases)  
Single insertions  
Single mutations



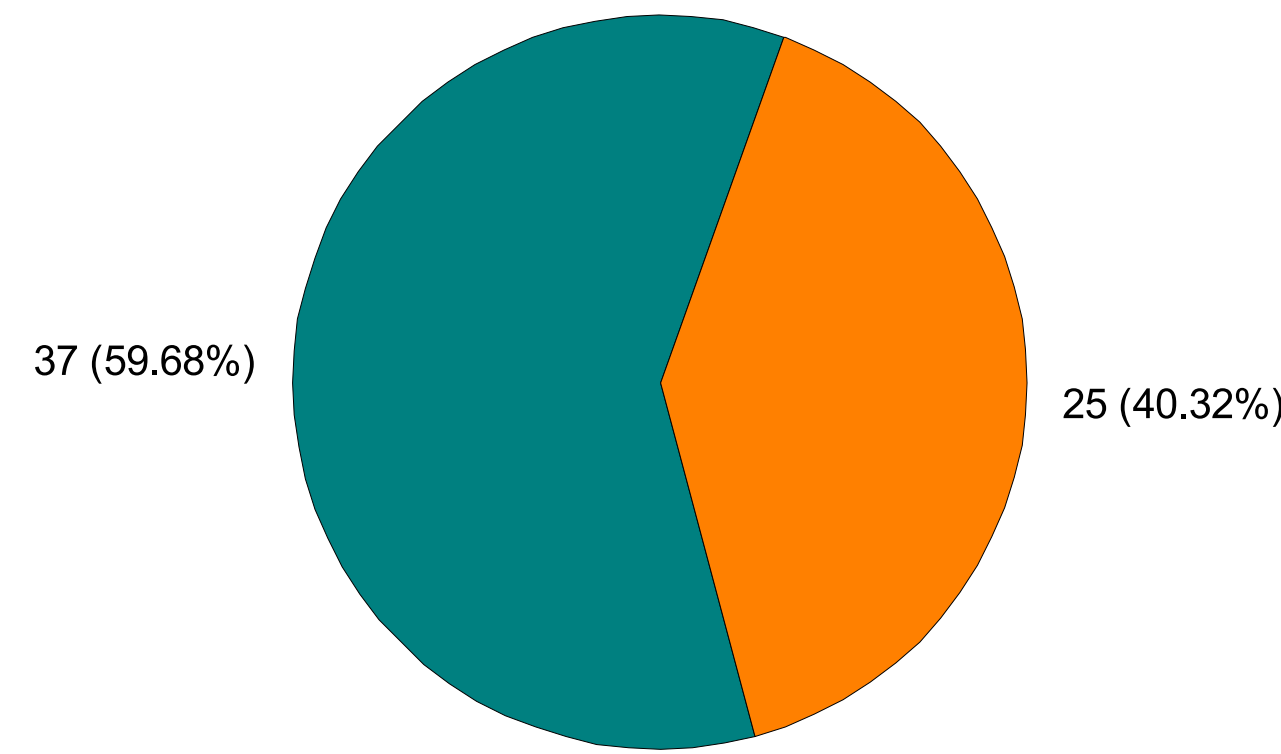
Single deletions  
Short deletions (2-4 bases)  
Long deletions (>4 bases)  
Single insertions  
Single mutations  
Ambiguous bases



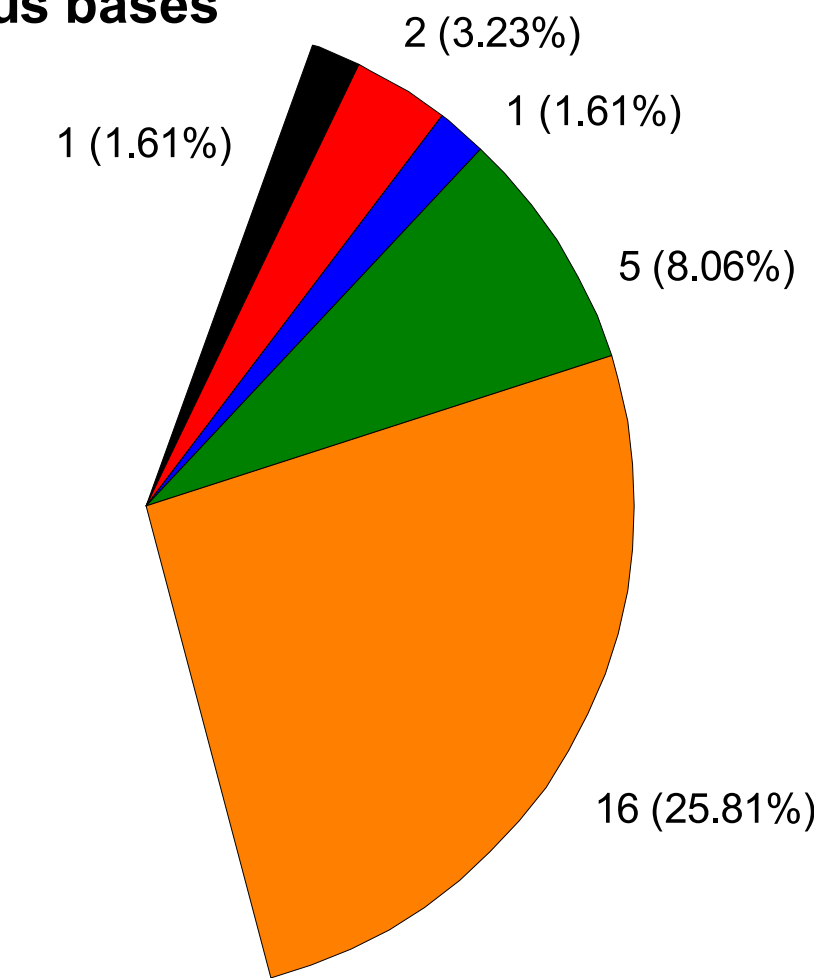
*E. coli* library

b.

Perfect promoters  
Promoters with errors

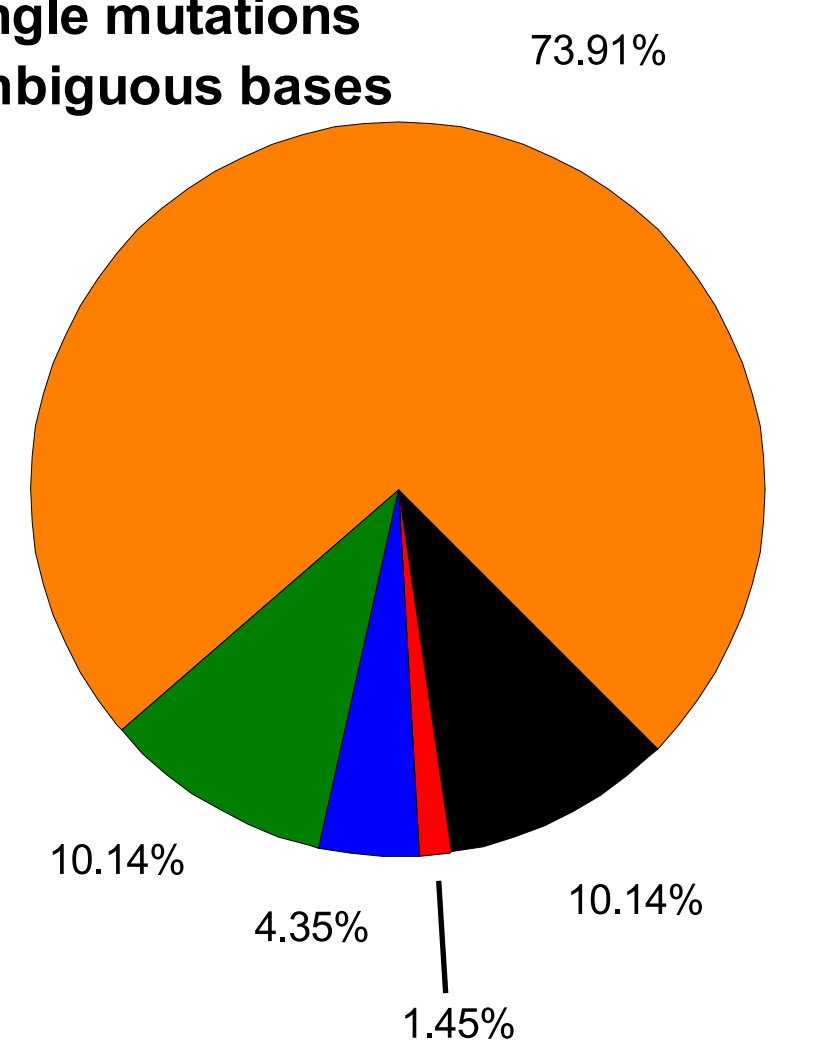


Single deletions  
Short deletions (2-4 bases)  
Single insertions  
Single mutations  
Ambiguous bases



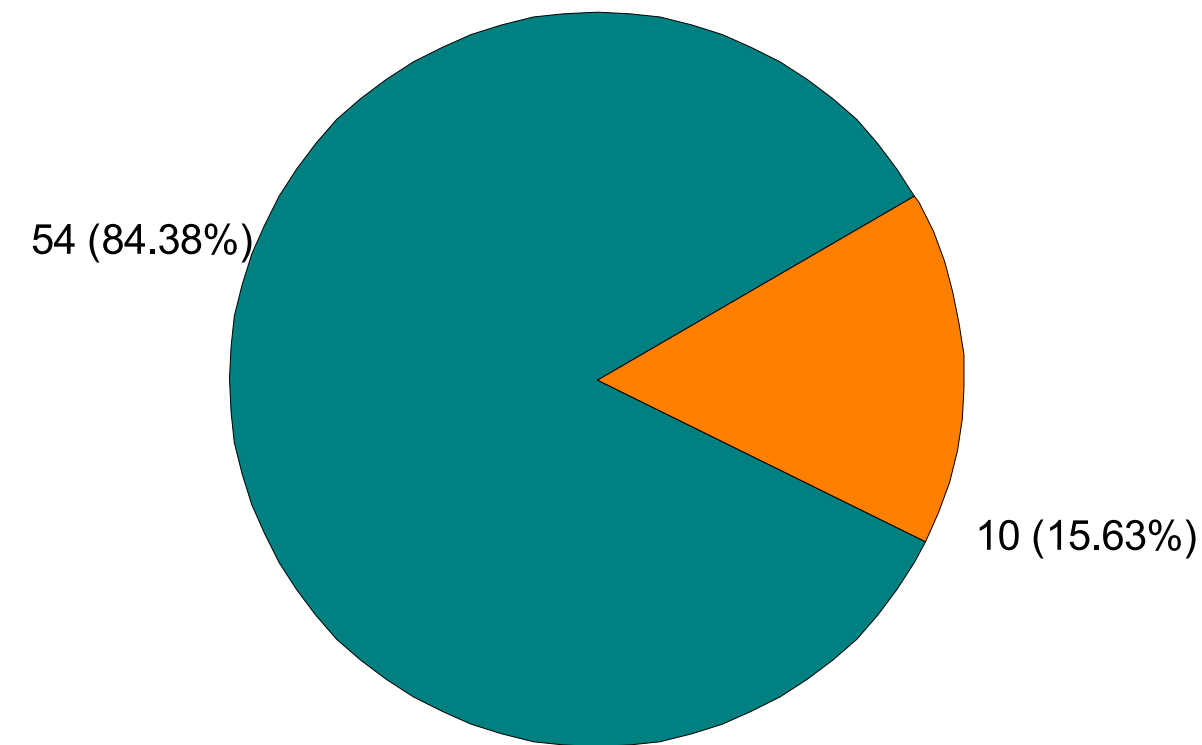
Yeast library

Single deletions (1 base)  
Short deletions (2-4 bases)  
Single insertions  
Single mutations  
Ambiguous bases

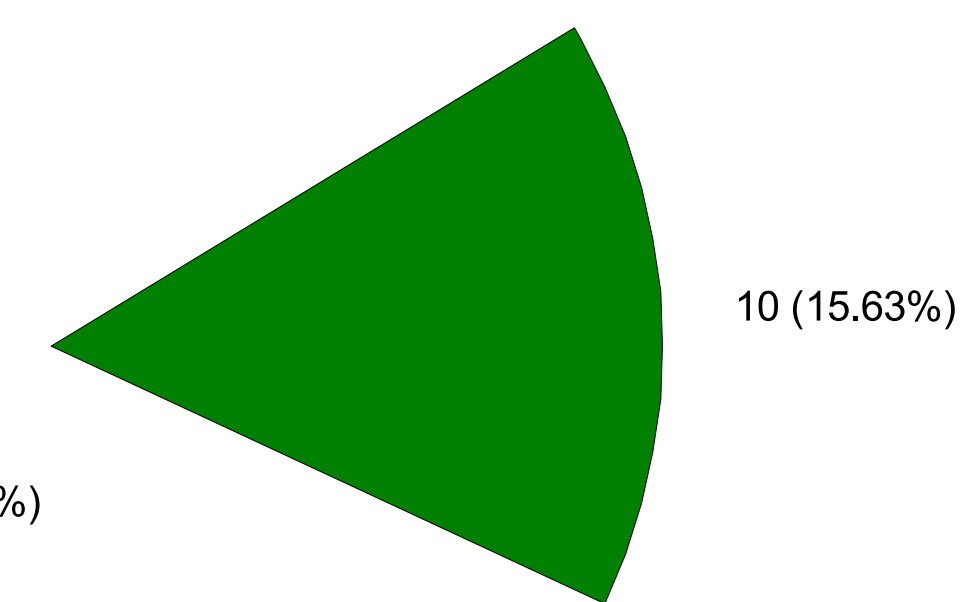


c.

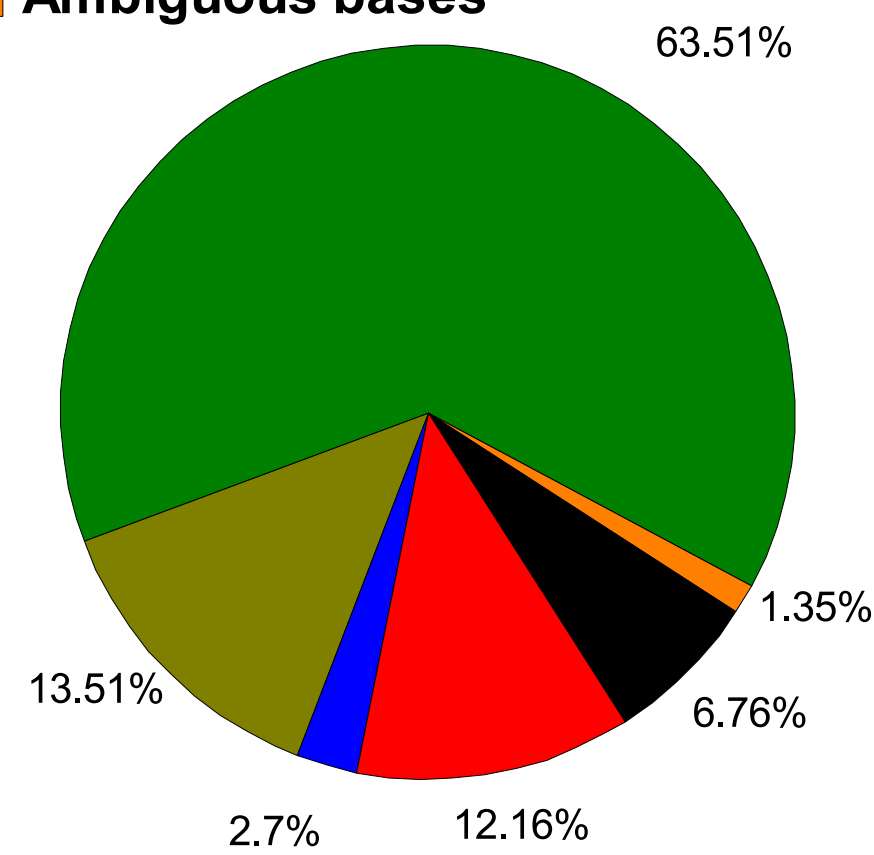
Perfect promoters  
Promoters with errors



Single deletions



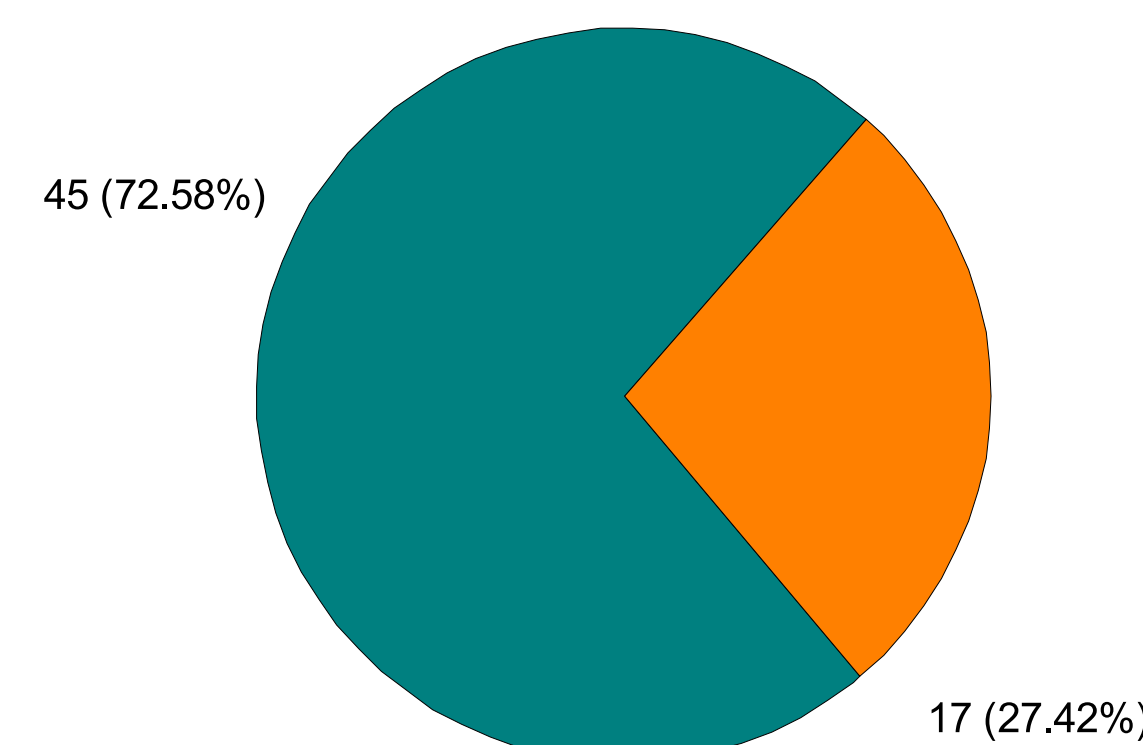
Single deletions  
Short deletions (2-4 bases)  
Long deletions (>4 bases)  
Single insertions  
Single mutations  
Ambiguous bases



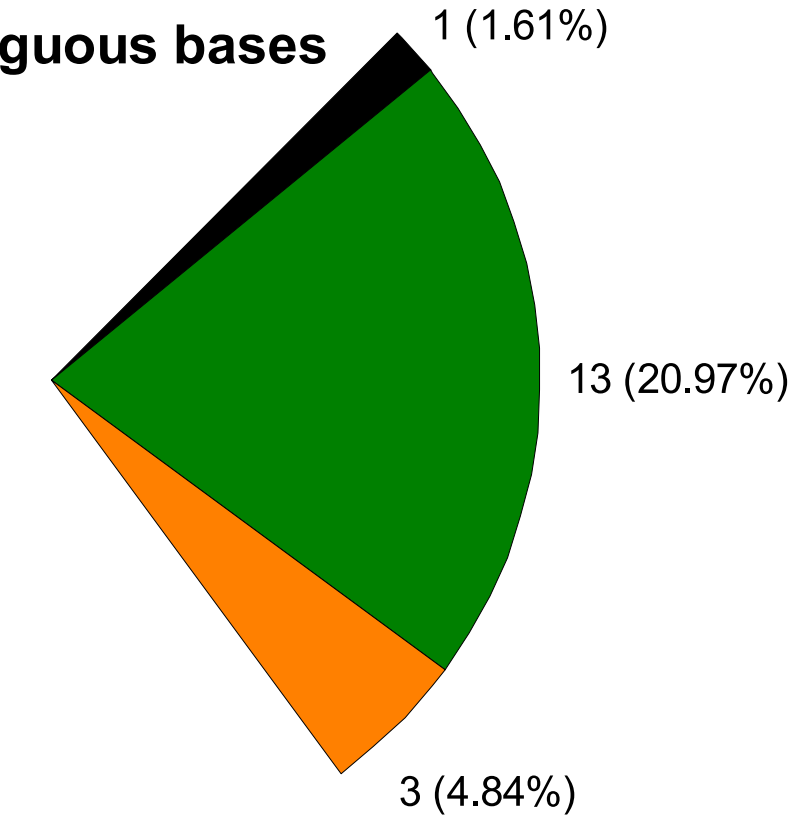
*E. coli* library

d.

Perfect promoters  
Promoters with errors

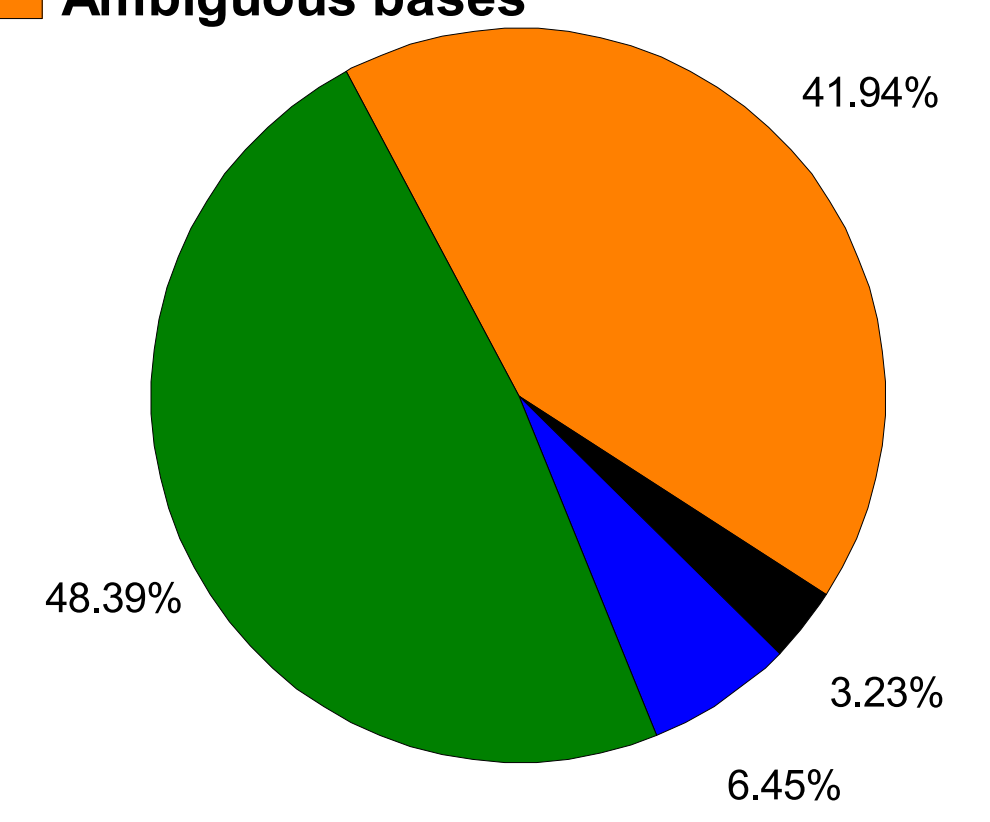


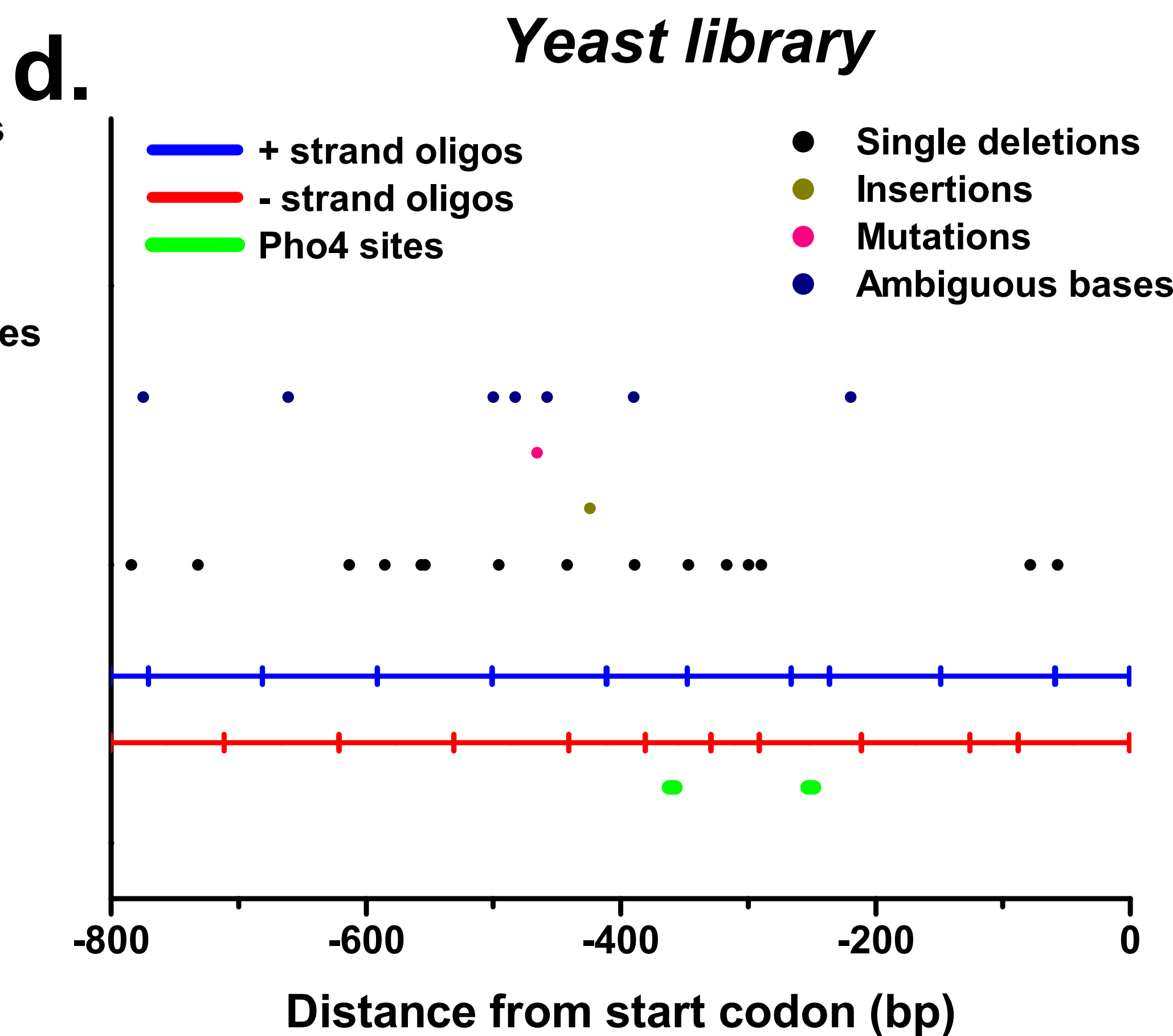
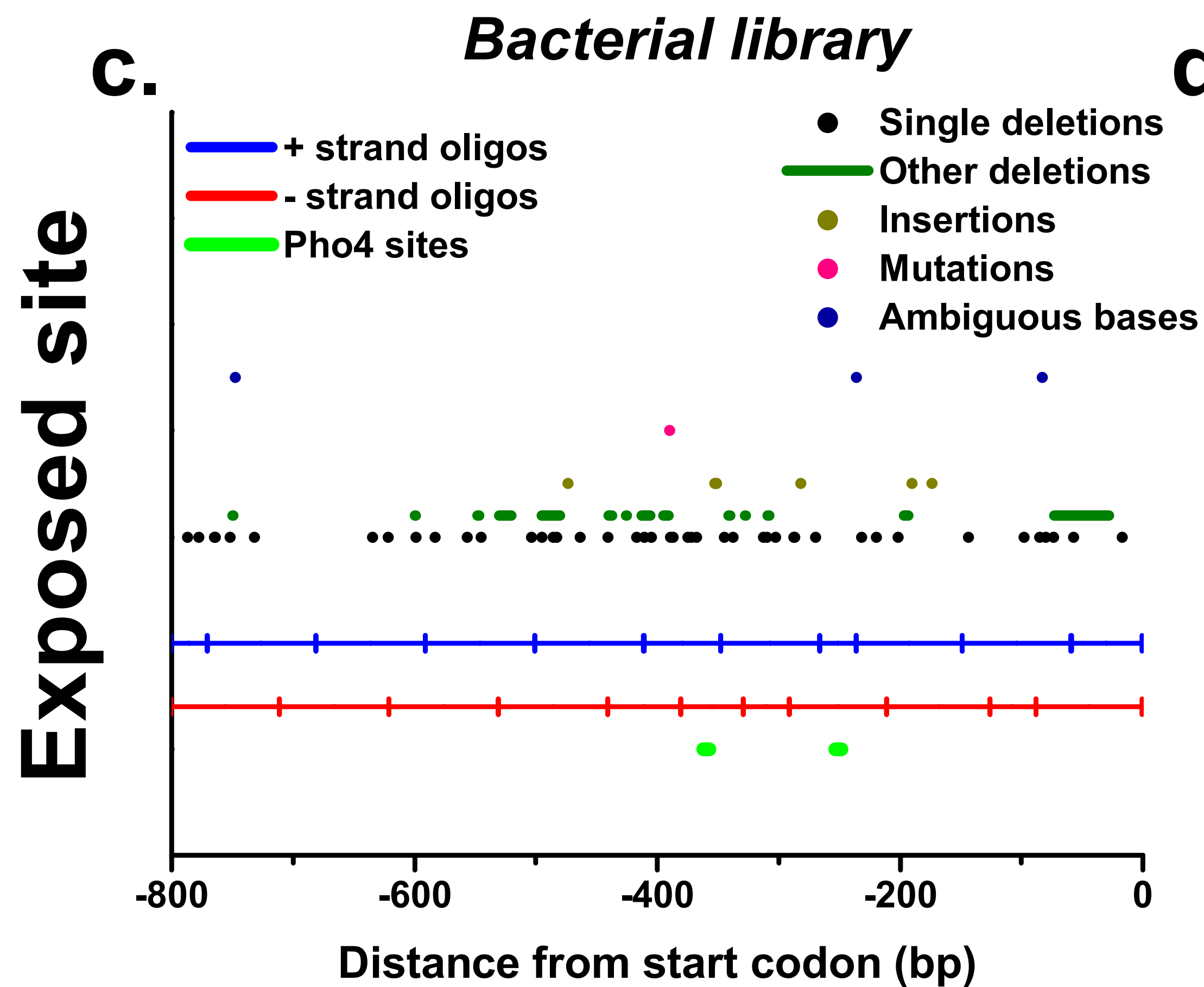
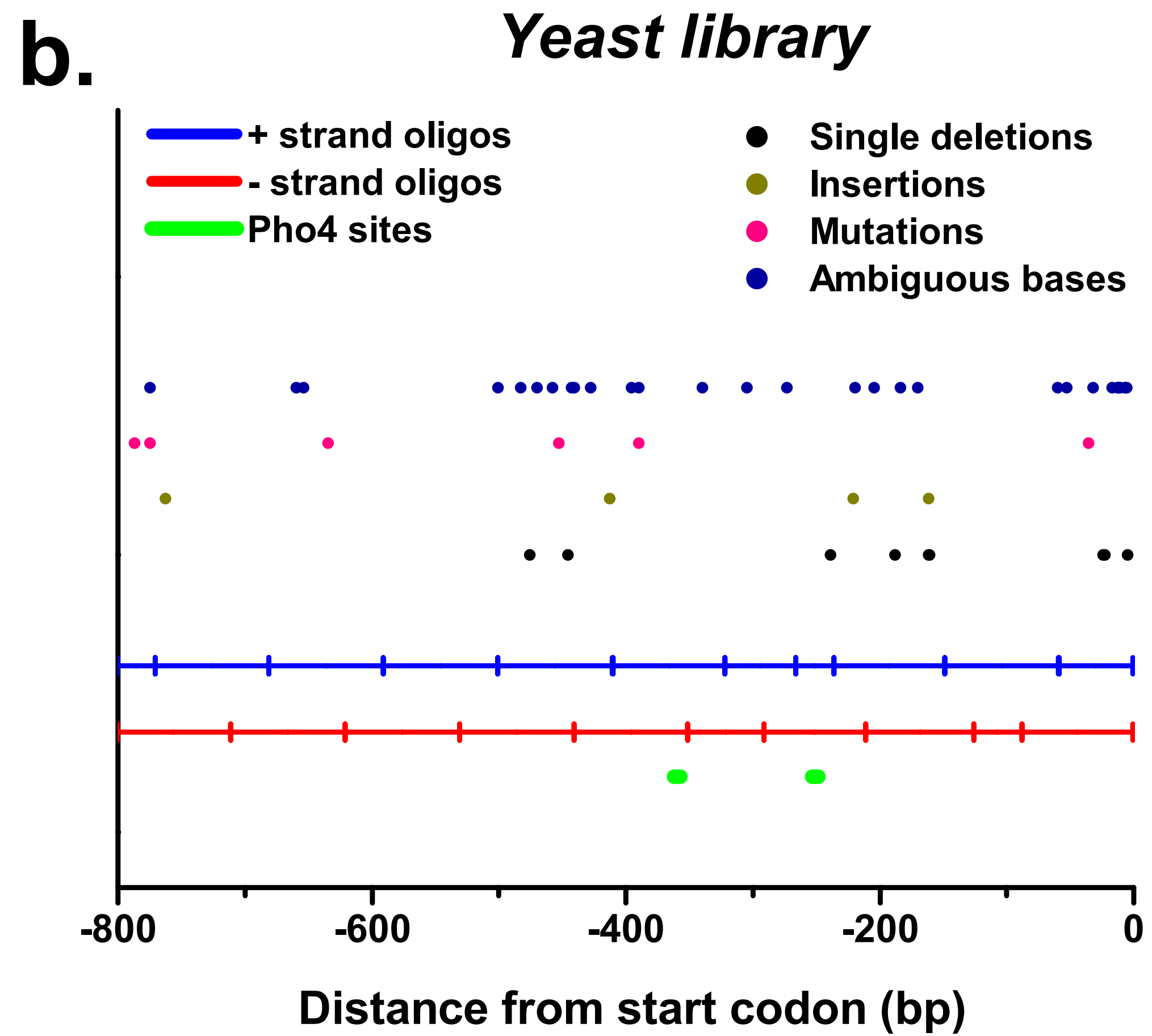
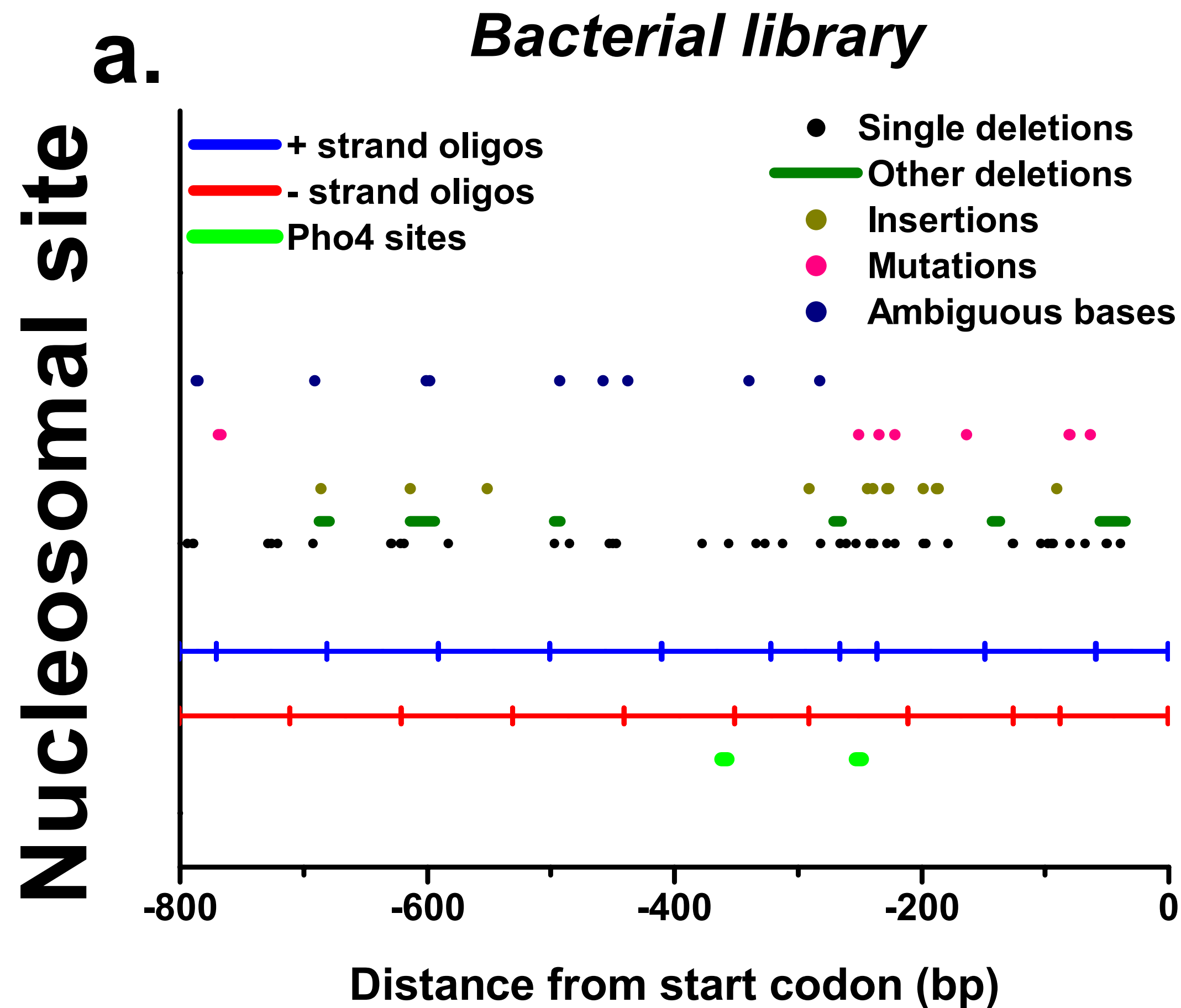
Single deletions  
Single mutations  
Ambiguous bases



Yeast library

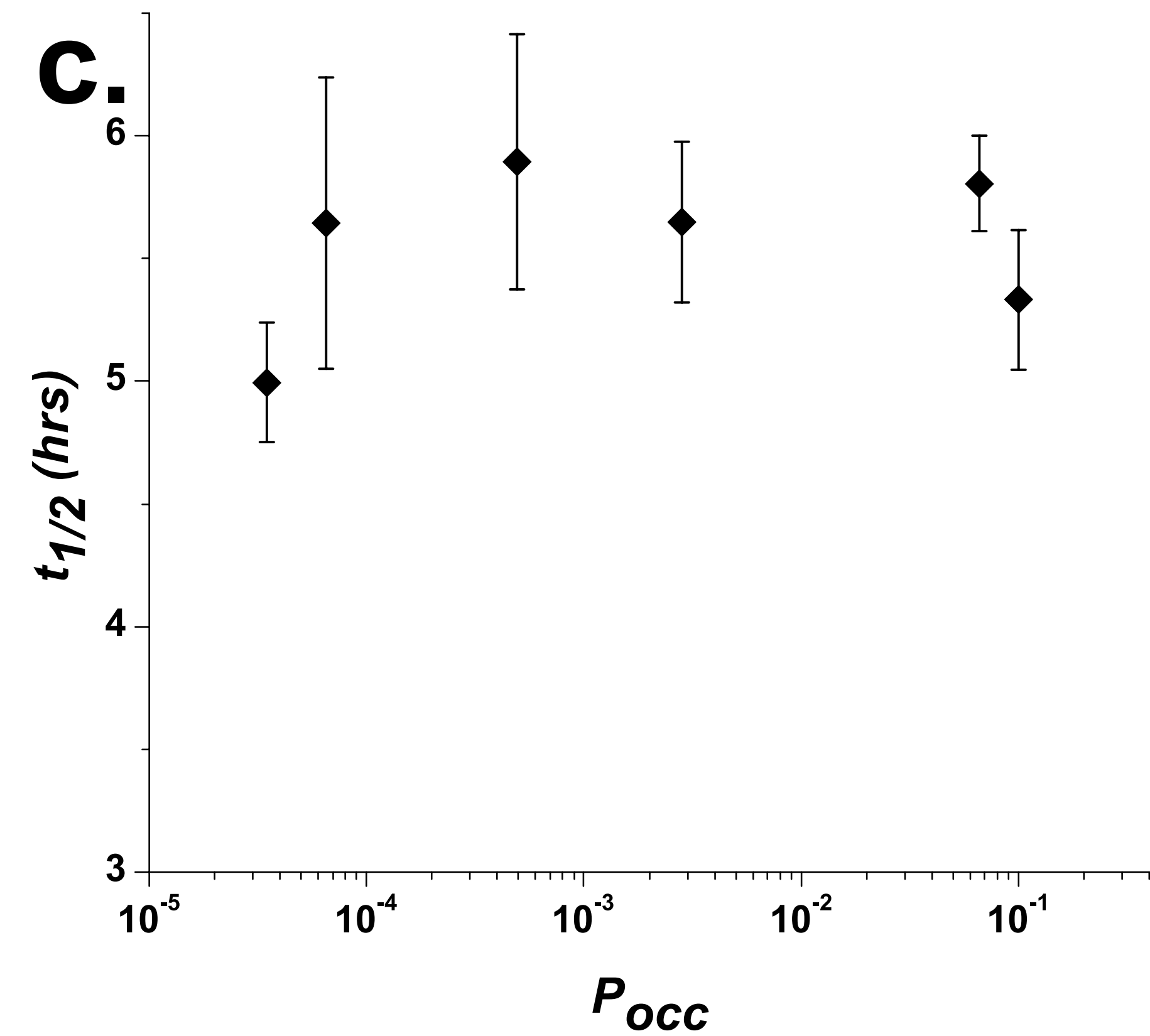
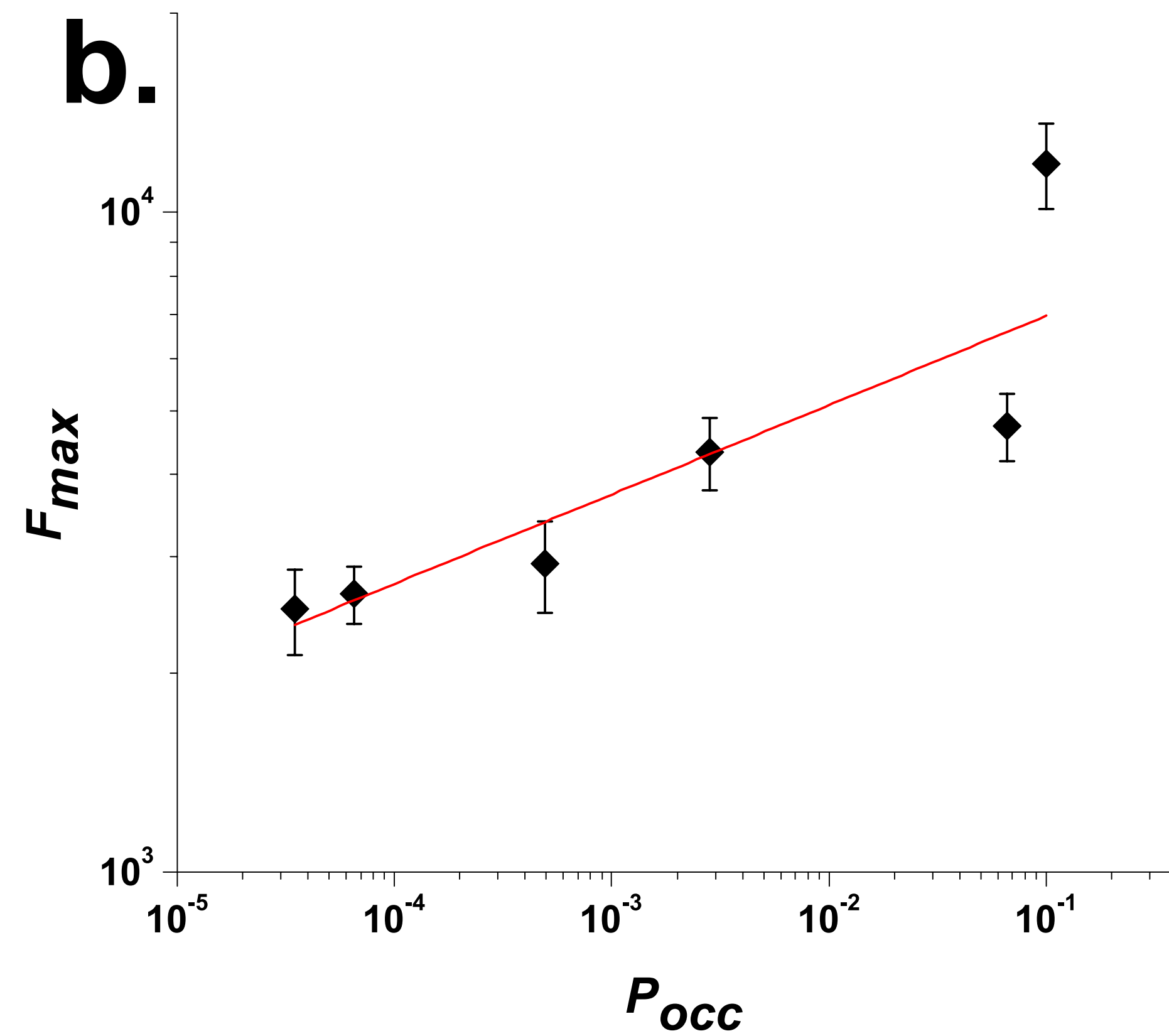
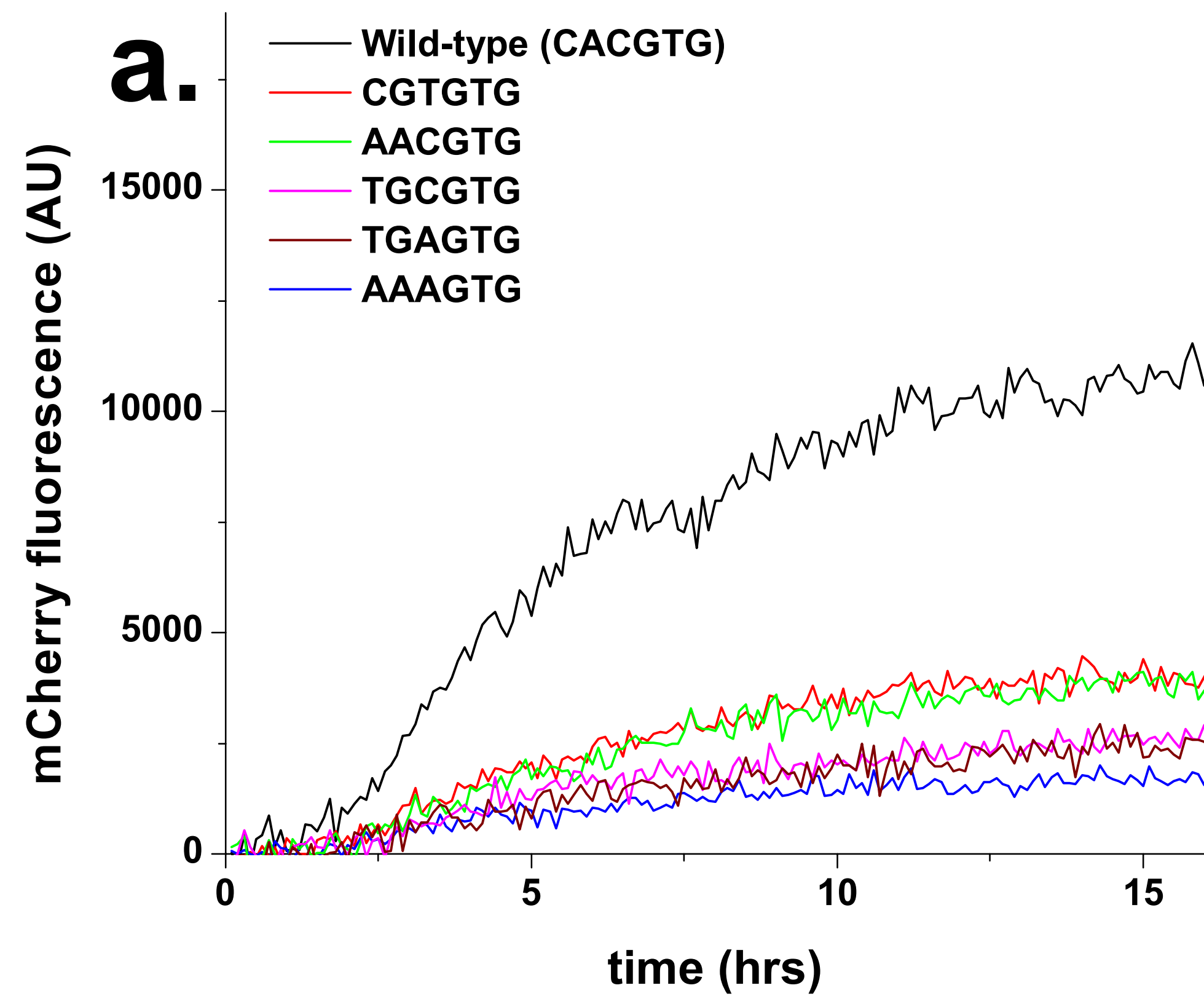
Single deletions  
Single insertions  
Single mutations  
Ambiguous bases



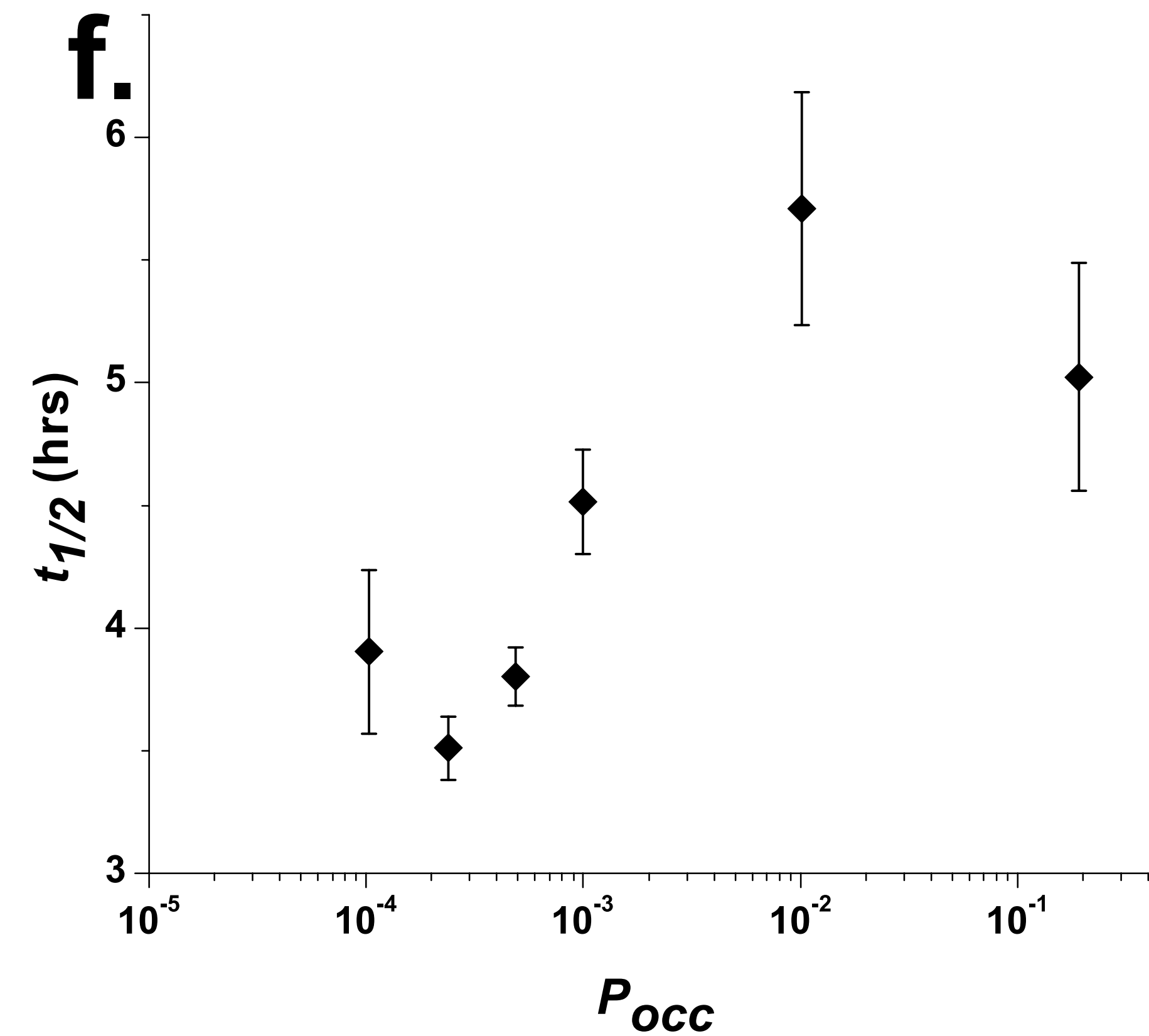
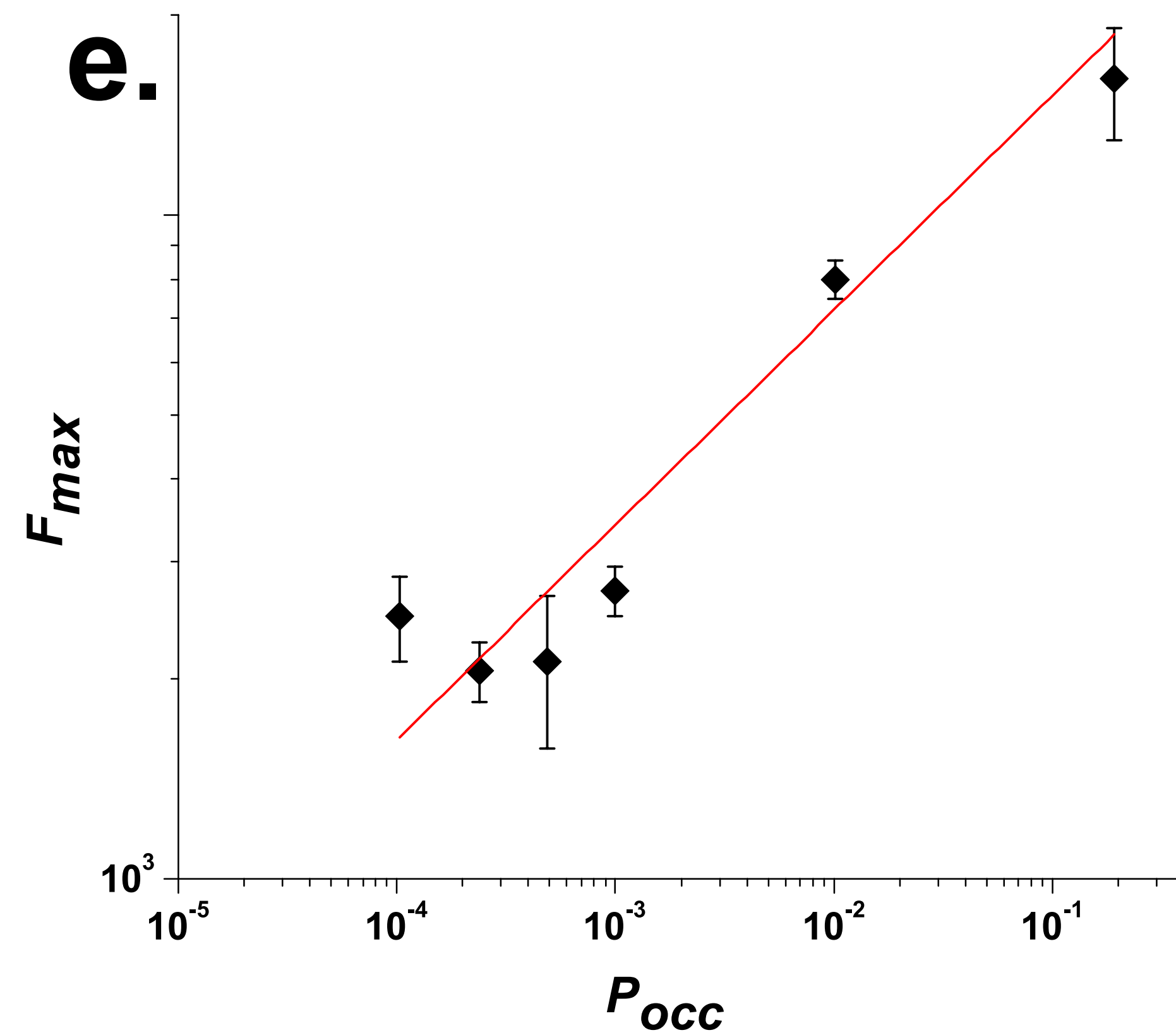
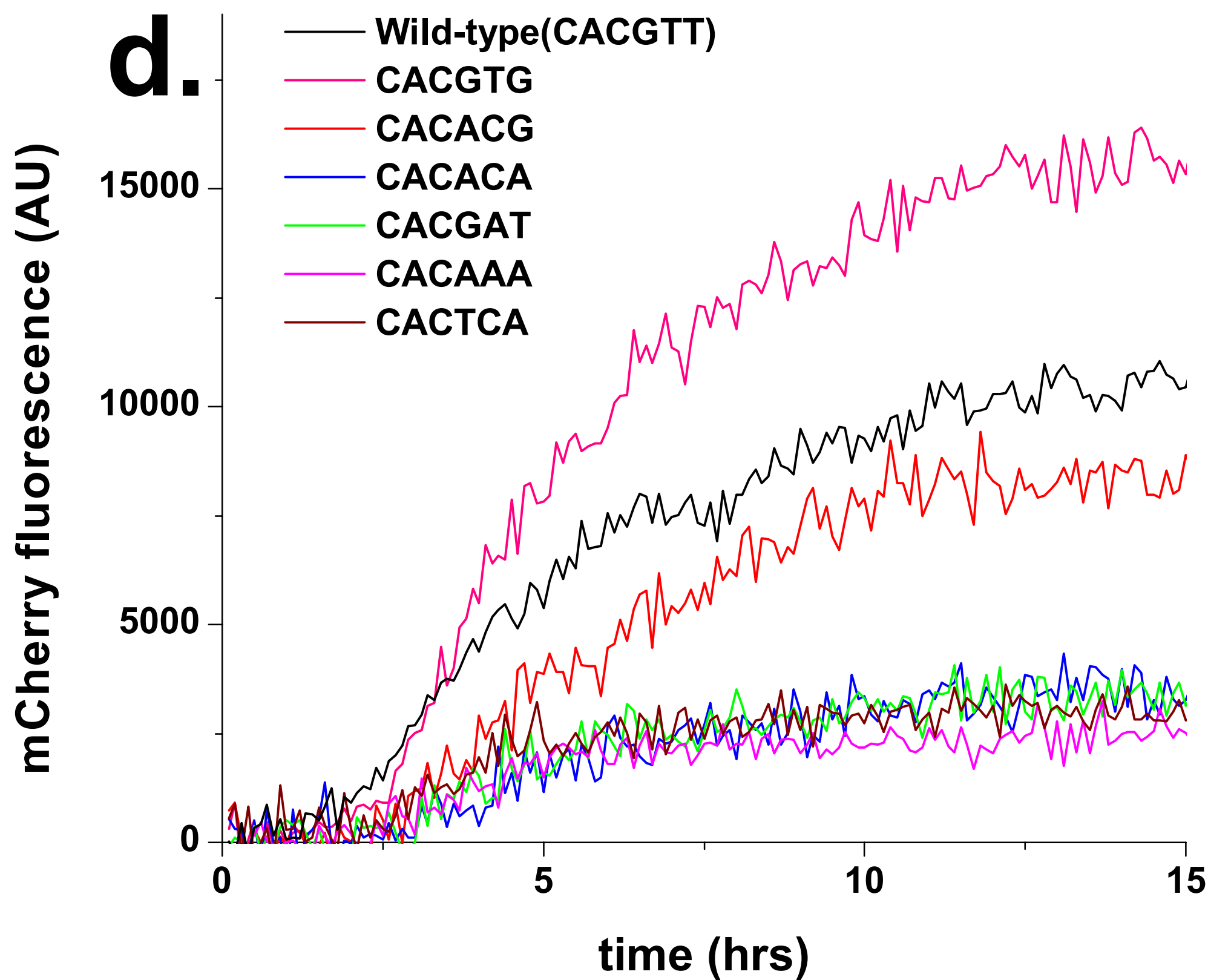


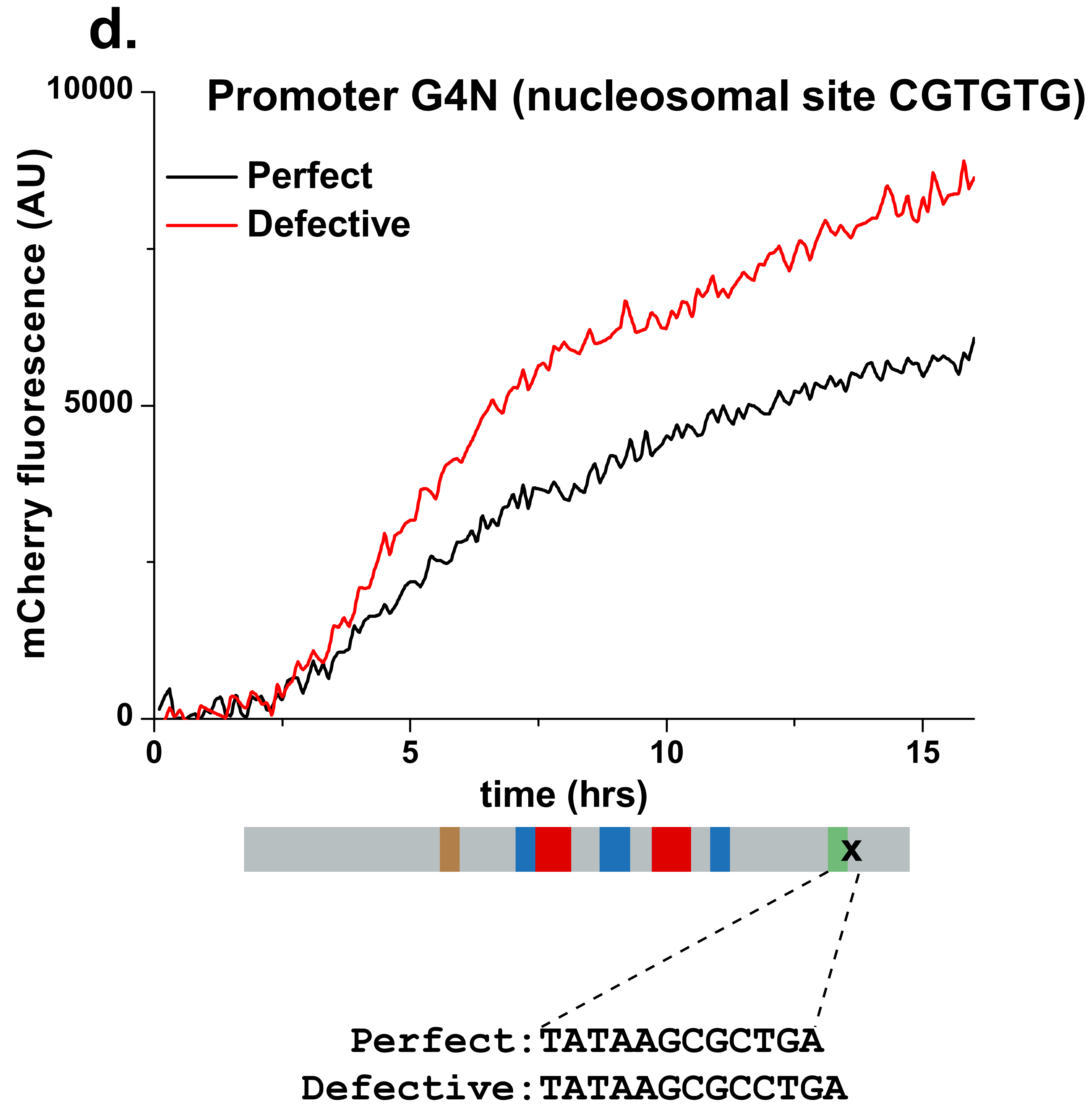
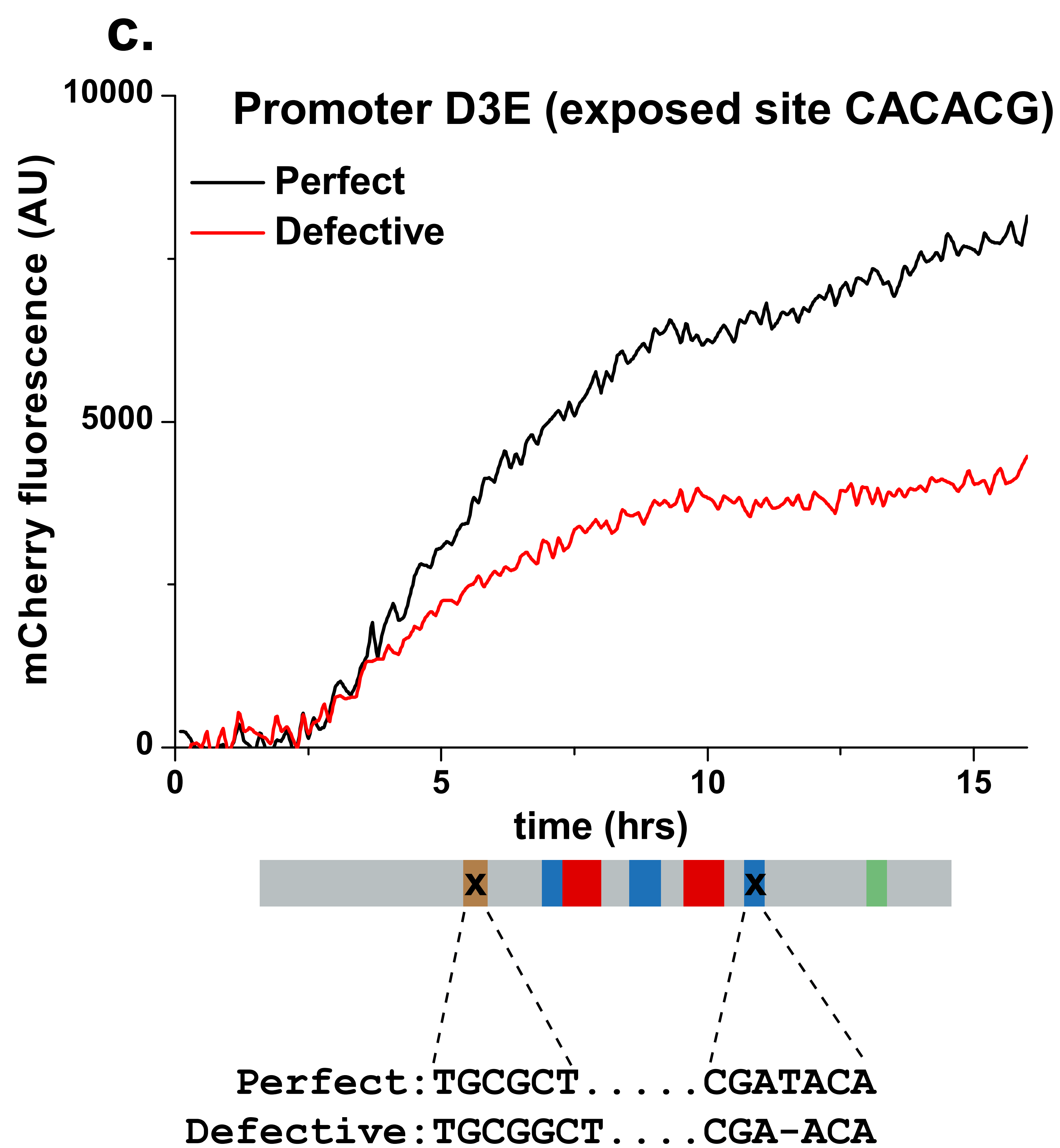
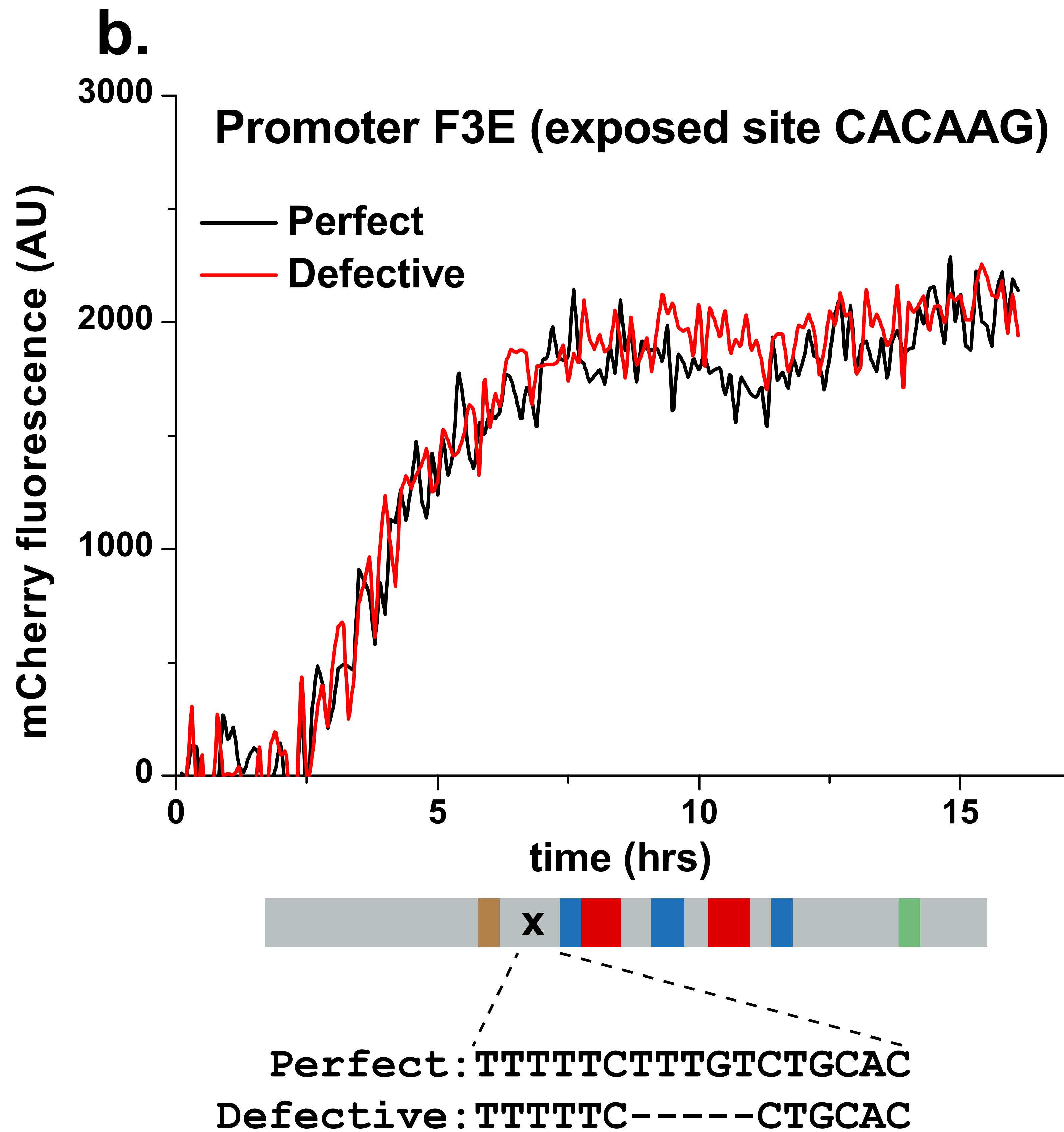
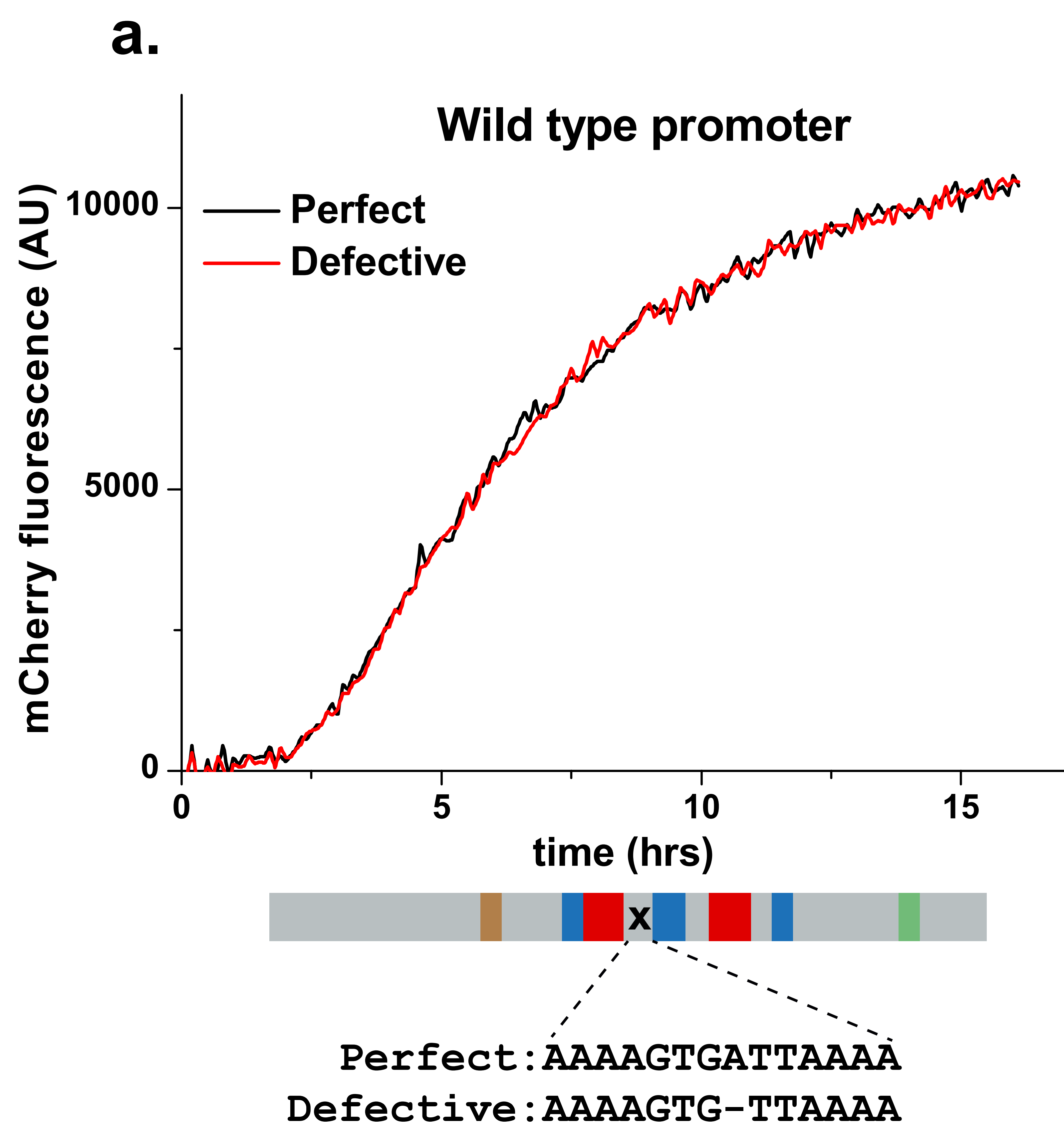


# Nucleosomal site



# Exposed site

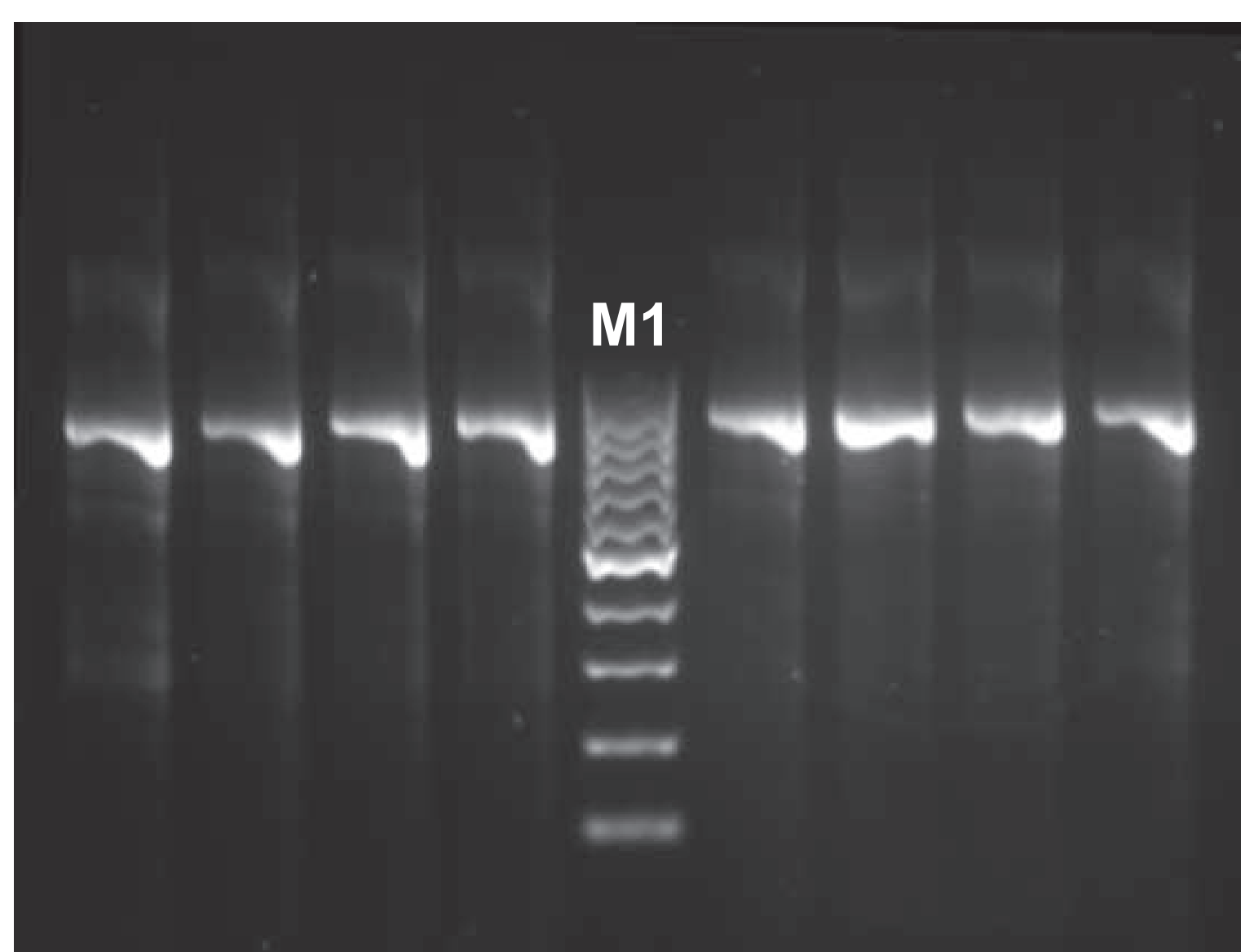




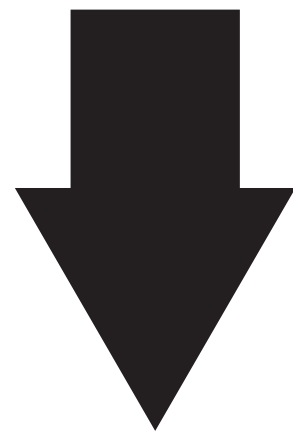
Pho4 site Pho2 site TATA box Swi5 site (predicted)



**a.**

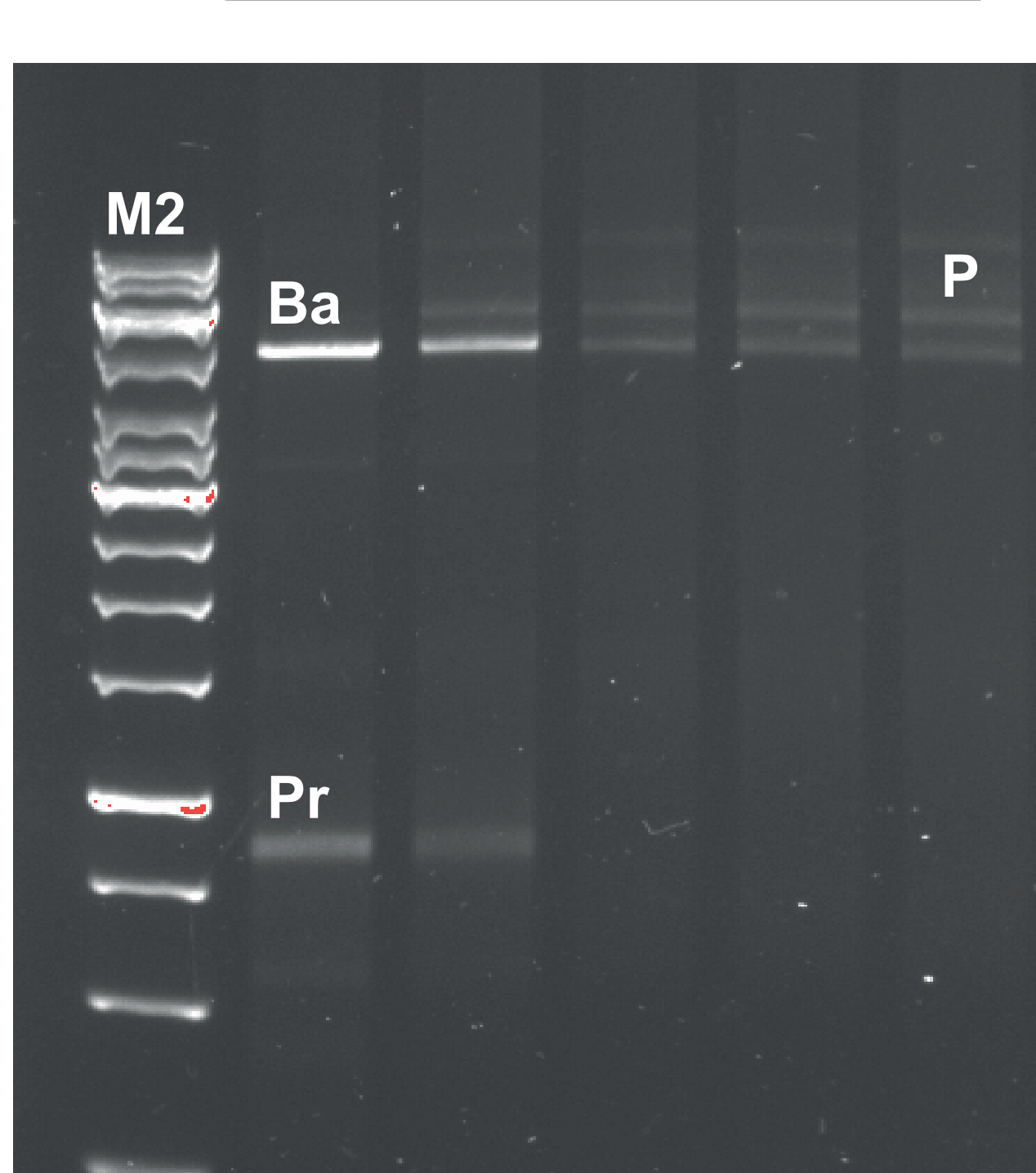


**Amplification of ligated promoters**

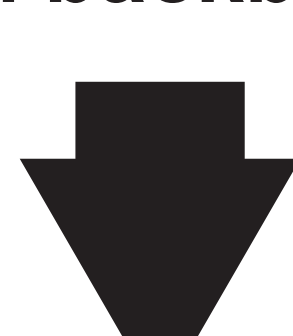


**b.**

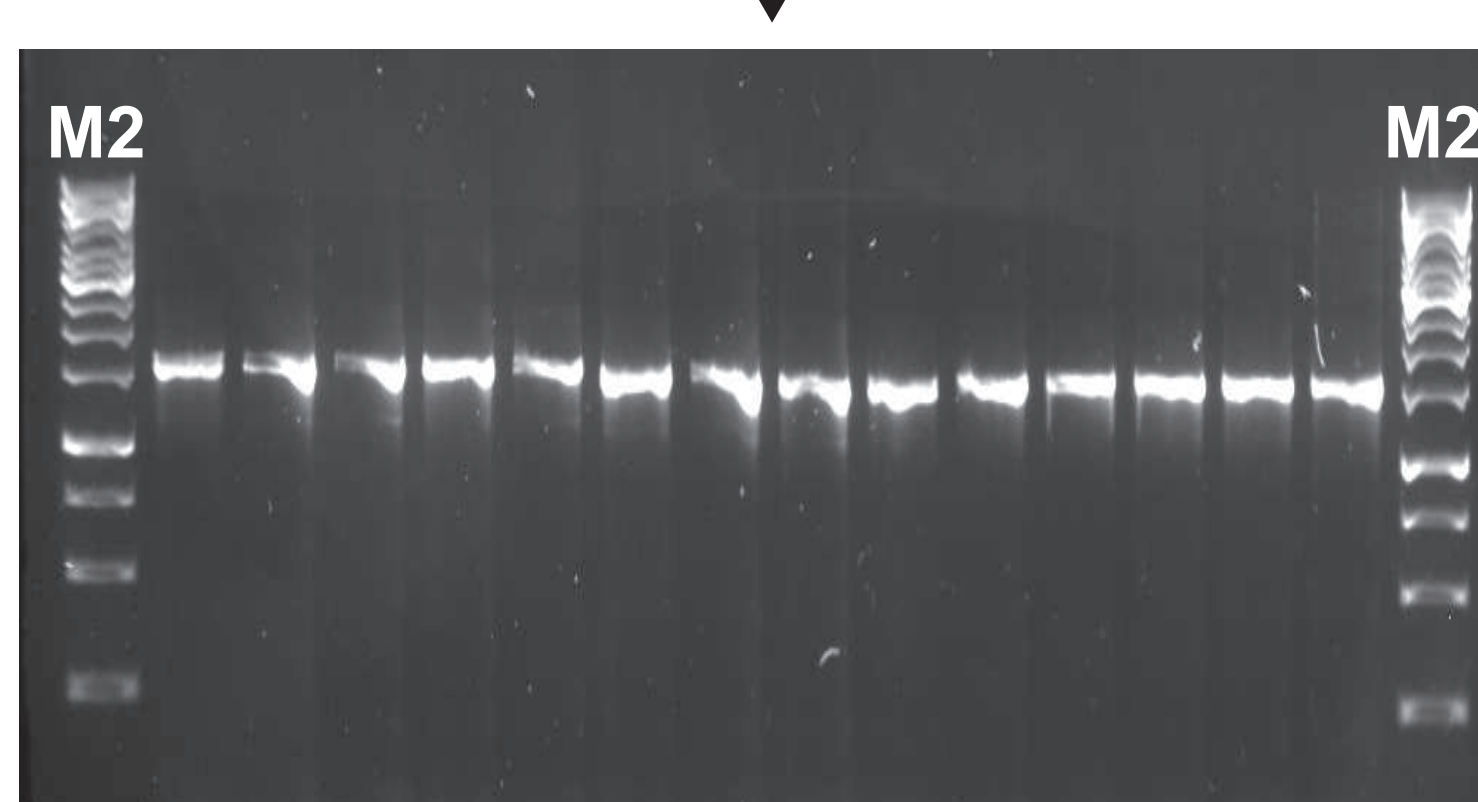
**Time**



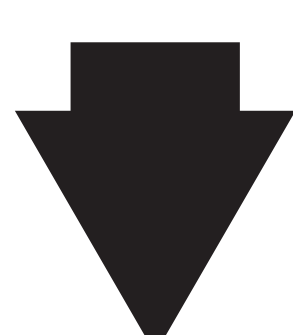
**Gibson assembly of promoter and plasmid backbone**



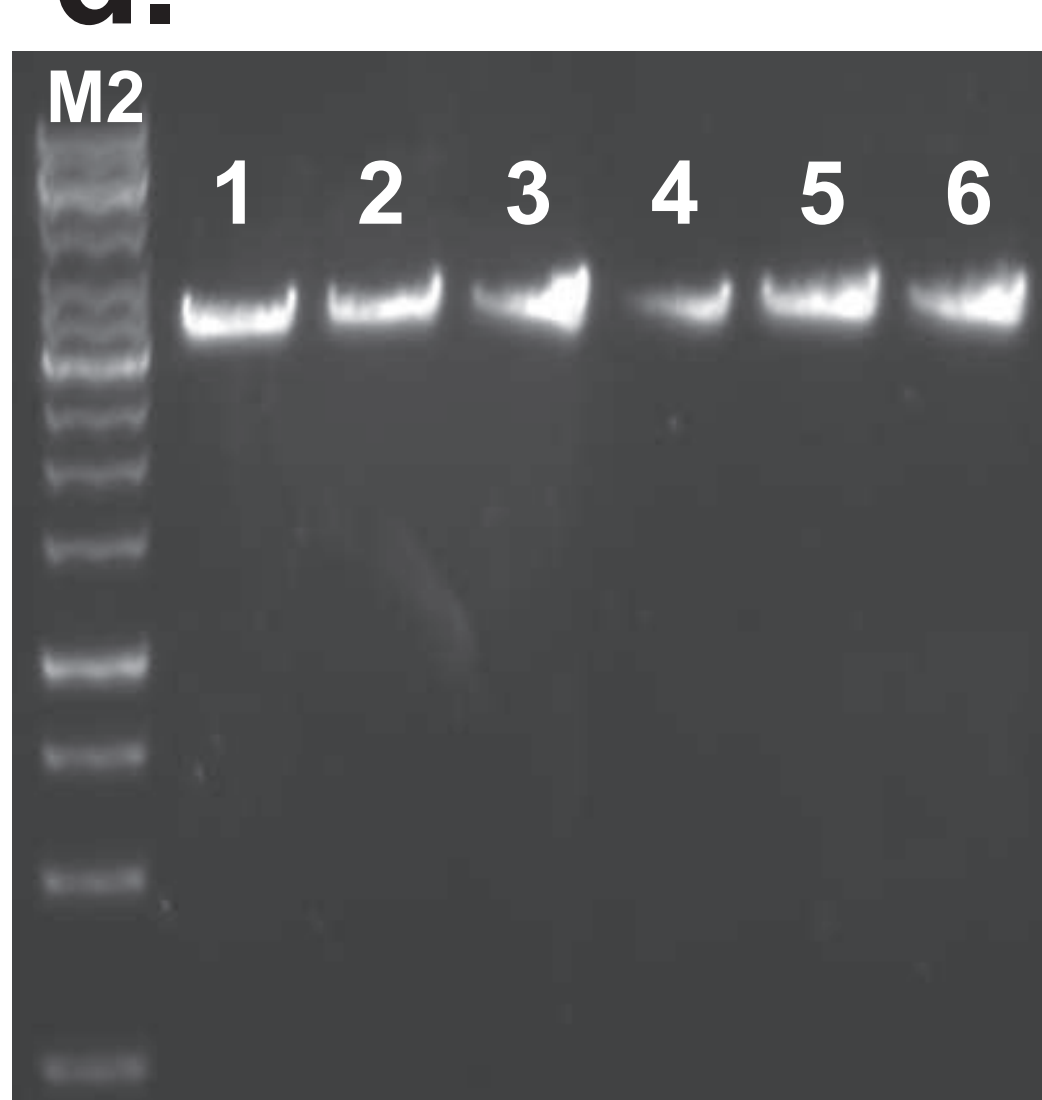
**c.**



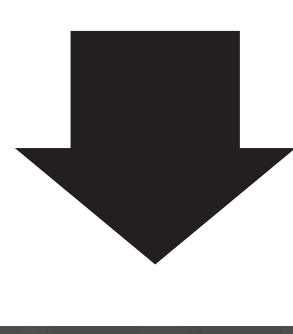
**PCR of promoters in plasmids from glycerol stocks of E.coli**



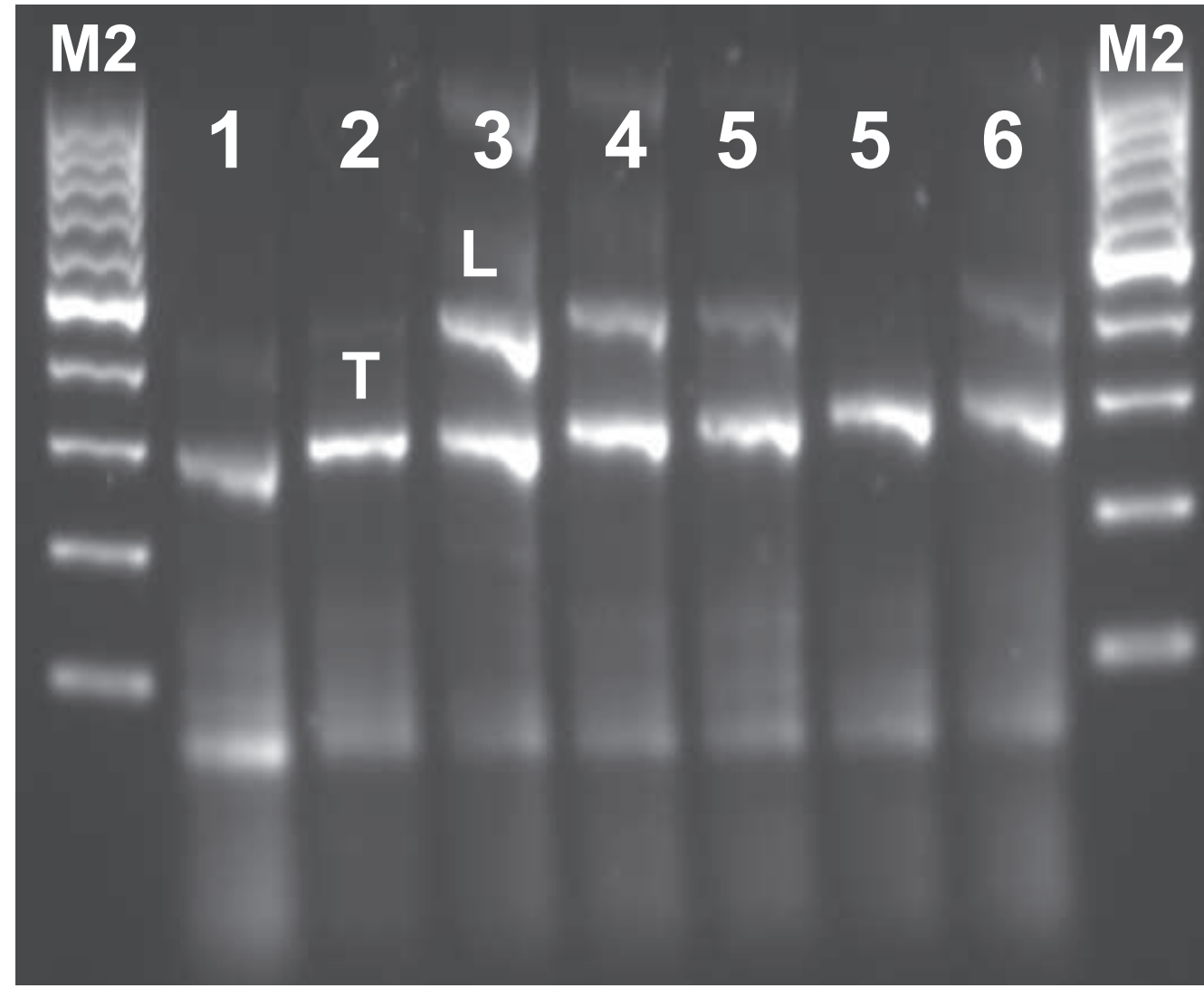
**d.**



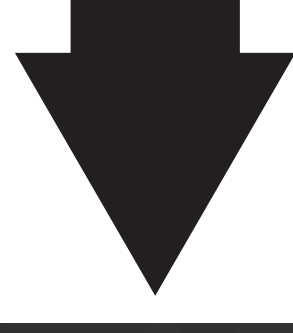
**Direct amplification of construct from plasmid**



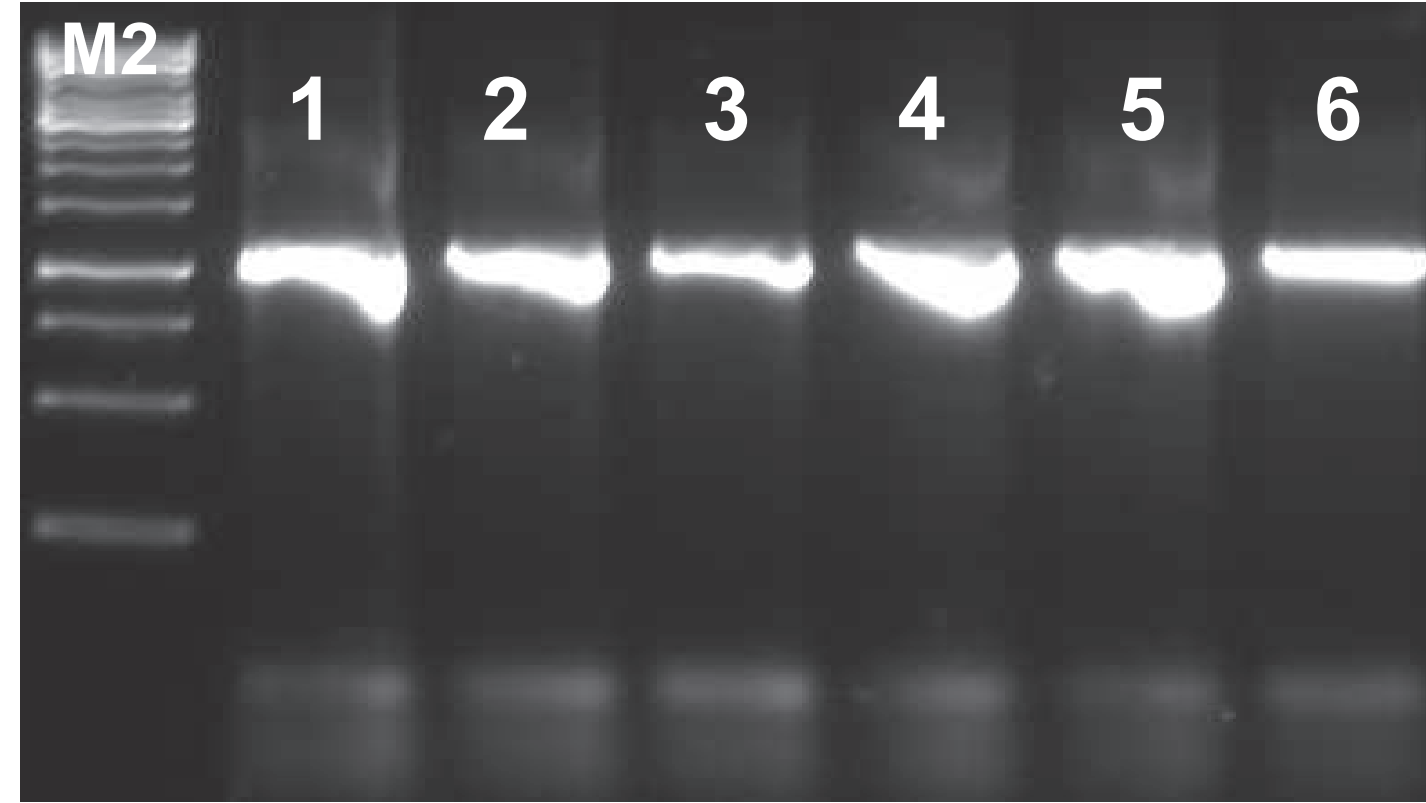
**e.**



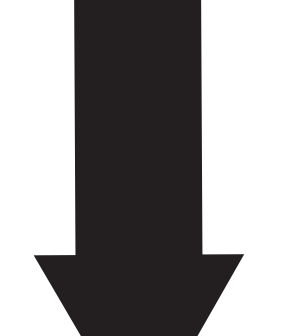
**Colony PCR of screened yeast transformants**



**f.**



**Synthetic promoter amplification from yeast**



**To sequence confirmation**